DISS. ETH NO. 23587

Second Generation Protein Replacement System to Study MLH1 Missense Mutations

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

Presented by

MARTIN FALKE
M.Sc. Ruprechts-Karls-Universität Heidelberg
18.02.1986

Citizen of Germany

Accepted on the recommendation of

Prof. Dr. Josef Jiricny
Dr. Mathias Gstaiger
Prof. Dr. Petr Cejka

2016

Für meinen Opa
Dr. Jürgen Krätzner

der mir in vielen Dingen
ein Vorbild war
und immer bleiben wird

Table of contents

1 Zusammenfassung

Hereditäres non-polypöses kolorektales Karzinom (HNPCC) oder Lynch-Syndrom ist eine vererbte genetische Erkrankung mit hohem Risiko für verschiedene Krebsarten, einschließlich kolorektales Karzinom, Endometrium- und Eierstockkarzinom. Die Ursache für HNPCC liegt in einem Defekt der DNA-Fehlpaarungsreparatur (englisch: DNA mismatch repair oder kurz MMR) durch Mutationen in *MMR*-Genen. Viele Studien untersuchten funktionell die verschiedenen Mutationen, die in Patienten entdeckt wurden und klassifizierten diese. Einige Varianten sind bis heute jedoch nicht eindeutig klassifiziert worden, wie z.B. die MutL homolog 1 (MLH1) Variante R659Q. Die Entwicklung neuartiger Methoden und Systeme erscheint notwendig, um auch diese Varianten verlässlich klassifizieren zu können.

Unser Labor hat vor kurzem das "all-in-one" Proteinersetzungssystem vorgestellt. Dabei wird das endogene Protein mithilfe von shRNA herunterreguliert und gleichzeitig eine Proteinvariante von cDNA induziert. Sowohl die shRNA als auch die cDNA sind auf einem einzigen Plasmid vorhanden und sind mit Doxycyclin induzierbar. Das System wurde erfolgreich verwendet, um die endogene DNA-Polymerase δ durch Mutanten zu ersetzen.

In dieser Arbeit haben wir das ursprüngliche pAIO-System durch einige Veränderungen verbessert und genutzt, um als Proof-of-Principle den Phänotyp der MLH1-Mutante R659Q zu studieren. Wir haben das Gateway-System in das Plasmid eingefügt, was die Generierung von verschiedenen Varianten extrem vereinfacht. Weiterhin haben wir eine FRT-Stelle (englisch: Flp Recombination Target) eingefügt, die die Herstellung von stabilen Zelllinien effizienter und einfacher macht. Ebenso wurden Möglichkeiten gefunden, den CMV-Promoter so zu verändern, dass die Expression der cDNA in verschiedenen Stärken möglich ist.

Anschließend haben wir Zelllinien mit GFP-MLH1 Wildtyp (WT), der MMR-defizienten Variante K84E und der Variante R659Q generiert und analysiert. Unsere Ergebnisse zeigen, dass die R659Q-Mutation zu reduzierten Proteinmengen sowohl im Zellkern als auch im Zytoplasma führt. Weiterhin konnten wir bestätigen, dass die R659Q-Mutation in MMR funktionell ist.

Allerdings konnten wir mithilfe von Pulldown-Experimenten zeigen, dass MLH1 R659Q im Gegensatz zum WT und K84E viel weniger an die Proteine Fanconi anemia group J protein (FANCJ) und FANCD2/FANCI-associated nuclease 1 (FAN1) bindet, die beide mit der Reparatur von DNA-Interstrang-Vernetzungen (englisch: DNA interstrand crosslinks oder kurz ICLs) assoziiert werden. Dieser Linie folgend konnten wir mit Zellvitalitätsexperimenten zeigen, dass beide MLH1-Mutanten resistenter gegenüber Mitomycin C (MMC), das DNA-Vernetzungen auslöst, sind und weniger stark in der G2/M-Phase gestoppt werden. Aufgrund unserer Resultate vermuten wir, dass nicht eine Einschränkung von MMR, sondern eine veränderte Reparatur von ICLs wegen eingeschränkter Bindung von MLH1 R659Q mit FANCJ und FAN1 zu Mutationen führt und damit zur Krebsentwicklung beiträgt.

Der zweite Teil dieser Arbeit beschäftigt sich mit der Charakterisierung der Proteininteraktionen von MMR. Obwohl das minimale System für MMR *in vitro* bekannt ist, bleibt noch zu klären welche Faktoren einen Einfluss auf MMR *in vivo* haben oder die Reparatur unterstützen. Ebenso ist offen welche Faktoren eine Rolle in nicht-kanonischer MMR spielen. Wir haben die MMR-Faktoren MLH1 und MutS homolog 2 (MSH2) mit GFP bzw. eGFP fusioniert und sie anschließend mit allen gebundenen Proteinen mithilfe von GFP-Trap angereichert und mit Flüssigchromatografie-Massenspektrometrie (englisch: liquid chromatography mass spectrometry oder kurz LC-MS) analysiert. Wir haben einige interessante Faktoren wie z.B., WD Repeat And HMG-Box DNA Binding Protein 1 (WDHD1) als Faktoren für MMR identifiziert, konnten aber mit dieser Methode keine Proteine speziell für nicht-kanonische MMR entdecken. Um dieses Hindernis zu überwinden, haben wir neuartige Techniken wie BioID und APEX ausprobiert und für unsere Forschungsfrage nutzbar gemacht. Beide Techniken basieren auf der Biotinylierung von endogenen Proteinen in unmittelbarer Nähe. Für beide Techniken wurden MMR-Fusionsproteine generiert und die Biotinylierung von endogenen Proteinen erfolgreich durchgeführt. Ebenso wurden Techniken für die Isolierung der biotinylierten Proteine wie auch für die anschließende Analyse mit LC-MS entwickelt.

2 Summary

Hereditary nonpolyposis colorectal cancer (HNPCC) or Lynch syndrome is an autosomal dominant genetic disease associated with a high risk of colorectal, endometrial and ovarian cancer. The cause of HNPCC is a defect in DNA mismatch repair (MMR) due to germline mutations in *MMR* genes. Cells defective in MMR show a mutator phenotype and microsatellite instability (MSI). Many studies have been conducted to characterize the phenotype of individual mutations in order to classify them and estimate the risk for the mutation carriers. Although most of the mutations have been classified as pathogenic or non-pathogenic, some remain uncertain such as the MutL homolog 1 (MLH1) R659Q mutant. To classify those variants, it is required to develop new techniques.

Our group previously reported on the development of an inducible protein replacement system, which we used for the phenotypic characterization of missense mutations in polymerase-δ. Our all-in-one plasmid (pAIO) inducibly expresses an shRNA that knocks down an endogenous mRNA and concurrently expresses an exogenous mRNA encoding the protein of interest or its variant. We now describe an improved pAIO plasmid with several new features. AttR sites for Gateway cloning and an FRT site (Flp-In system) for faster generation of stable cell lines were introduced. In addition, strategies to alter the CMV promoter to obtain strong, medium, low and very low cDNA expression were developed.

We took advantage of the second generation pAIO plasmid to study phenotypic variants of the MMR factor MLH1. We generated stable Flp-In T-REx-293 cell lines expressing GFP-MLH1 wild type (wt), an MMR-deficient K84E mutant and an R659Q mutant. Subsequently, we induced our system, which resulted in the replacement of endogenous MLH1 with GFP-MLH1. We found that protein levels of GFP-MLH1 K84E were more abundant and MLH1 R659Q levels were reduced in whole cell extracts. The K84E mutant, but not the R659Q mutant, showed MMR deficiency *in vitro* and *in vivo*. Pulldown experiments with GFP-Trap revealed that the binding of the R659Q mutant to DNA repair factors Fanconi anemia group J protein (FANCJ) and FANCD2/FANCI-associated nuclease 1

(FAN1) was diminished. Hence, we studied cell viability of the variants after treatment with DNA interstrand crosslink (ICL) inducing agents and found that the K84E mutant as well as the R659Q mutant showed resistance against mitomycin C (MMC). In addition, both mutants showed reduced G2/M phase arrest after MMC treatment. These results imply erroneous MLH1-dependent ICL repair in MLH1 R659Q cells, which provides a potential explanation of how MLH1 R659Q carriers acquire mutations leading to tumor development.
Together, our experiments demonstrate that the improved pAIO system is generally suitable to study genotype-phenotype correlations and might help to elucidate a mechanism underlying disease development for specific mutations. With the help of our system, we could show that the MLH1 R695Q variant does not have a defect in MMR but that its impaired binding to FANCJ and FAN1 might result in alternative, probably mutagenic, ICL repair, which leads to tumor development.

The second part of this thesis focuses on the characterization of the MMR interactome under different aspects of DNA metabolism. Although the mechanism of MMR *in vitro* is largely known, we still require more information about how MMR is facilitated *in vivo* to explain the whole mechanism in the setting of a living cell. Moreover, additional information about the MMR interactome is needed to better understand non-canonical MMR, which describes a process that works independently of DNA replication. It can be activated outside of S phase by alkylating agents and was shown to be mutagenic. GFP and eGFP-tagged MMR proteins were generated and stable cell lines were produced using the pAIO system. GFP-Trap pulldown followed by liquid chromatography mass spectrometry (LC-MS) helped identify some interesting proteins including WD Repeat And HMG-Box DNA Binding Protein 1 (WDHD1). Unfortunately, the approach was not suitable to identify specific factors for non-canonical MMR. Hence, novel techniques such as BioID and APEX, which are based on proximity-dependent biotinylation, were explored and tested for the study of the MMR interactome after DNA damage. MLH1 BioID and APEX fusion proteins were cloned and stable cell lines generated. Moreover, we developed

strategies to efficiently pull down biotinylated proteins and analyze them by LC-MS.

3 Introduction

3.1 General aspects of cancer development

Cancer is described as a group of diseases that have one thing in common: All types of cancer derive from normal tissue and proliferate without control. They have the potential to leave their initial site and spread all over the body in a process known as metastases formation, which is the main cause of death among patients. Cancers are often referred to as tumors, although not all tumors are cancerous. In general, tumors are divided into two groups: benign tumors that do not metastasize and malignant tumors (cancers) that do metastasize. A lot of time and resources have been spent to find the origins of cancer and many substances called carcinogens have been associated with tumor development (1). They are divided into three categories: physical carcinogens such as ultra violet and ionizing radiation, chemical carcinogens including tobacco smoke or substances produced by living organisms, e.g. aflatoxin, and oncogenic viruses. In principle, they all cause cancer by producing genotoxic stress, which leads to mutations in the genome (2). These mutations can either activate pro-proliferative genes called oncogenes or inactivate anti-proliferative genes called tumor suppressors. Both events can transform normal tissue cells into cancer cells that grow in an uncontrolled manner and can metastasize in the body and lead to the disease we call cancer (1).

Not all cancers, however, are caused by carcinogens. 5-10% of all cancers are caused by inherited mutations. This phenomenon is called familial cancer. Mutations in tumor suppressor genes involved in cell proliferation such as *restinoblastoma* (*RB*) and *adenomatous polyposis coli* (*APC*) or DNA repair such as *breast cancer 1* (*BRCA1*) and *mutL homolog 1* (*MLH1*) are causes of familial cancer (1,3). In most cases, only one allele is mutated and the normal, wild type (wt) allele is affected over time by another mutation or loss of heterozygosity (LOH), which is consistent with the "two-hit hypothesis" by Knudson (1971), which states that both alleles of a tumor suppressor need to be inactivated before a cell transforms (4). Once a cell has lost its remaining functional copy, key mechanisms such as cell growth, proliferation, apoptosis control or DNA

repair are significantly impaired and the chance of this cell to transform into a cancer cell increases dramatically (1,2).

3.1.1 Mutations and loss of heterozygosity – the second hit on the road to cancer

Usually, familial cancer patients inherit only one mutated allele and the second is wt, though rare cases with homozygosity for the mutant allele have been described such as inherited biallelic mismatch repair deficiency. In such a scenario, both parents need to be carriers. Then, the patients develop tumors very early and have dramatically decreased life expectancy (5). Heterozygous carriers are diagnosed much later, which can be explained by the fact that the wt allele gives raise to a functional protein that is dominant over the mutant. Obviously, the wt allele needs to be inactivated in order to acquire the phenotype, which can lead to a cancerous cell, such as higher mutation rates because of impaired DNA repair pathways. The functional copy can be lost or inactivated by different mechanisms, including deleterious point mutations, deletions, or hypermethylation of the promoter region known as epigenetic silencing. Hypermethylation of the promoter region, which shuts down transcription, has been observed for many tumor suppressor genes. *MLH1* in particular is frequently inactivated by epigenetic silencing (6). LOH can occur through chromosome missegregation or mitotic recombination. In the process of chromosome missegregation, a heterogeneous cell loses its functional allele due to a failure of the spindle-assembly checkpoint (see Figure 1 A). Figure 1 B illustrates how mitotic recombination between homologous chromatids can cause LOH (7).

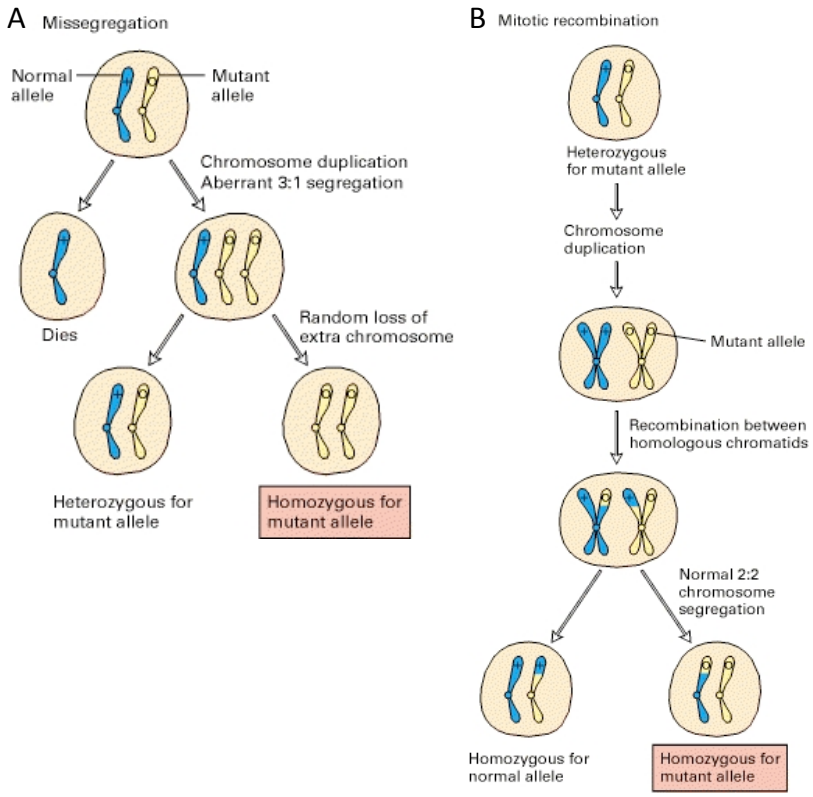


Figure 1: Mechanisms of LOH

(A**)** A heterozygous cell can lose the functional copy of a gene by failure of the spindle-assembly checkpoint. In such a scenario, the two daughter cells receive the alleles not in a 2:2, but in a 3:1 ratio. In the next cell cycle, one daughter cell can lose the third allele with the functional copy. **(**B**)** In the process of mitotic recombination, an exchange between homologous chromosomes after duplication of the genome results in two homozygous daughter cells: one with two wt alleles and the other with two mutant alleles (modified from Lodish et al., 2004).

3.2 Hereditary non-polyposis colon cancer

Hereditary non-polyposis colon cancer (HNPCC) or Lynch syndrome is a prominent example of familial cancer. It accounts for about 4% of all colorectal cancers (CRC) and is characterized by an inherited dominant predisposition to early onset of colon cancer as well as extracolonic cancers including endometrial, ovary, stomach and small intestine cancer. Cancers in the typical sites are usually diagnosed before the age of 50. The cumulative risk for CRC in HNPCC families is about 80% and the lifetime risk for females to develop endometrial cancer amounts to about 60% (8). At 12%, the risk for ovarian cancer is also much

higher compared to the general population with a risk of about 1.3% (see Figure 2) (9,10).

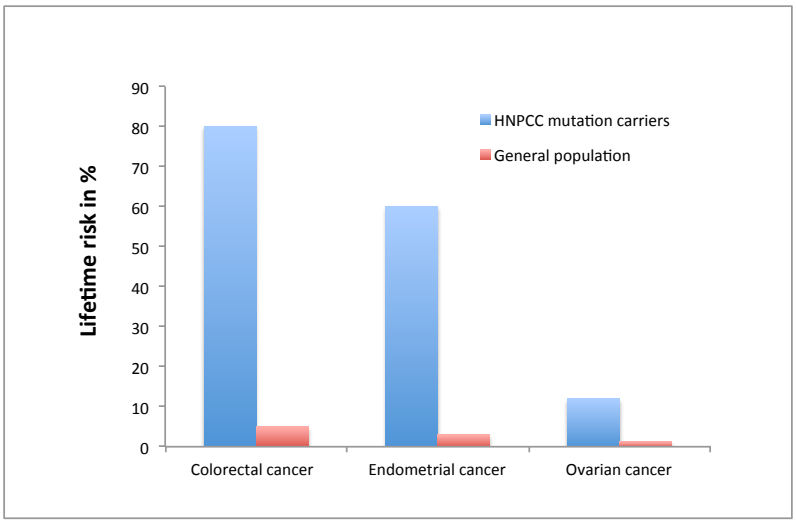


Figure 2: Lifetime risk of HNPCC mutation carriers and general population

HNPCC is caused by a defect in DNA mismatch repair (MMR), which recognizes and repairs errors arising during DNA replication and recombination (11,12). MMR-deficient cells acquire point mutations, insertions and deletions in coding and non-coding sequences and are thus hypermutagenic, which eventually leads to tumor development. DNA polymerases display increased error rates when they replicate repetitive motifs known as microsatellites, such as $(A)_n$ or $(CA)_n$. If these are left unrepaired by MMR, shortening of those areas can be observed – a phenomenon known as DNA microsatellite instability (MSI)(13). MSI is typically used for diagnosis of HNPCC (14).

Many studies have been conducted to identify HNPCC-predisposing genes and four MMR genes were found to be frequently mutated: *MLH1, PMS2, MSH2* and *MSH6*. In addition, *MSH3* and *PMS1* mutations have been associated with the syndrome, but need to be further examined (15). According to the InSight (International Society for Gastrointestinal Hereditary Tumors) database, about 48% of all germline mutations in HNPCC affect the *MLH1* gene. 37% of the mutations occur in *MSH2*, 11% in *MSH6* and less than 4% in *PMS2* (16).

3.2.1 Diagnosis, treatment and prognosis of HNPCC

HNPCC is described as a cancer syndrome with an early onset of CRC and other tumors caused by mutations that impair MMR. Although this description is rather comprehensible, it turned out to be not that simple in clinical practice. Specific guidelines were needed to clearly separate HNPCC from sporadic cancers also caused by mutations in MMR genes and other cancer syndromes such as familial adenomatous polyposis (FAP). The International Collaborative Group for HNPCC (ICG-HNPCC) introduced standardized criteria for HNPCC diagnostics in 1991. According to these criteria, known as the Amsterdam Criteria, a candidate needs to fulfill all of the following: three or more family members with a confirmed CRC (Amsterdam Criteria I) or HNPCC-related cancers (Amsterdam Criteria II) and one should be a first-degree relative of the other two, one or more cancers need to be diagnosed before the age of 50 years and FAP should be excluded (17,18). As an alternative, the National Cancer Institute introduced the Bethesda guidelines in 1997, which are similar to the Amsterdam Criteria but also include patients with HNPCC-related cancers under the age of 60, but with MSI (19). Once a family has been identified, the affected persons undergo genetic screening to search for mutations in *MMR* genes. Other family members are also screened for these mutations and carriers are recommended to join early-detection programs that include physical exam, colonoscopy and gynecologic methods for endometrial and ovarian cancer. Prognosis for HNPCC depends on the stage of the tumor at diagnosis. Although HNPCC-related CRCs in general are poorly differentiated, which normally correlates with poor prognosis, they have a better outcome than comparable sporadic cancers. To date, surgery is still the most effective treatment for HNPCC (14). Adjuvant chemotherapy does not improve the prognosis, but seems to slightly decrease the outcome in patients with high MSI (20).

3.3 Defects in DNA mismatch repair as a source of genome instability

When a cell divides, its genome must be duplicated with as few errors as possible. A large set of proteins described as the replication machinery takes part in this process known as DNA replication. The actual copying of the DNA strands is carried out by a specific group of enzymes called DNA polymerases. Several DNA polymerases with widely different fidelities have evolved in cells. The error rates range from 10^{-1} to 10^{-8} falsely incorporated bases per nucleotide for low-fidelity and high-fidelity polymerases, respectively (21). In addition to base substitutions, insertion or deletion loops (IDLs) that occur due to incorrect primer annealing during DNA replication or due to polymerase slippage especially at microsatellites, are a source of replication errors (see Figure 3) (13,22). Moreover, replication errors can be induced by environmental stress, such as high oxygen radical levels, DNA-methylating agents or radiation. Under normal conditions, the replication error rate is estimated to be around 10^{-7} to 10^{-8} (21). The post-replicative MMR system that recognizes spontaneous base substitutions and IDLs increases replication fidelity by 50 to 1,000-fold, which results in a total error rate of about 10^{-9} to $^{-10}$ (23). A defect of the MMR system therefore results in increased genome instability due to higher mutation rates accompanied by MSI, which eventually leads to cancer development.

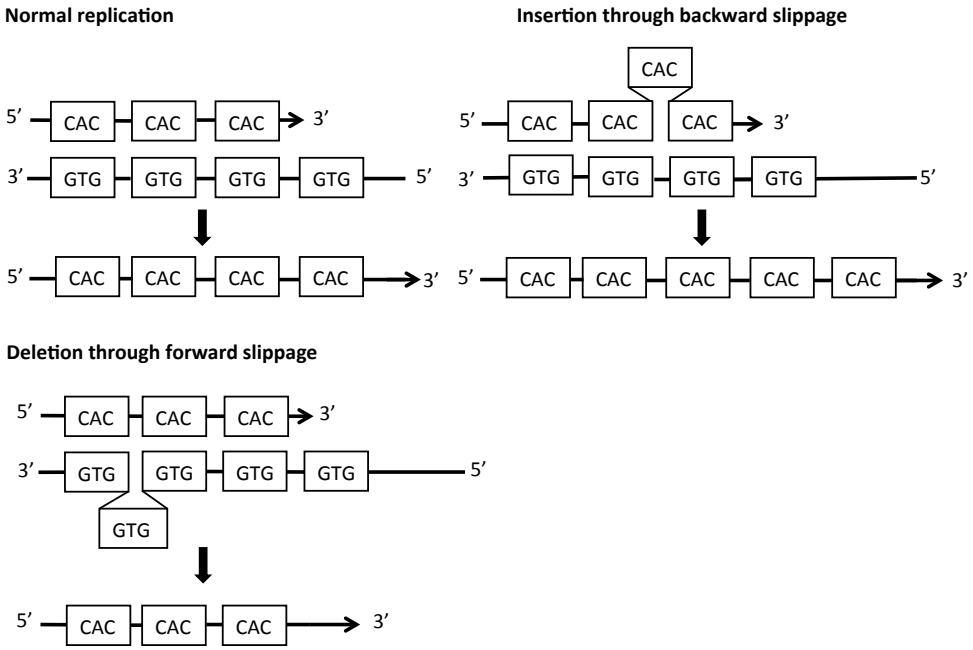


Figure 3: Formation of insertion and deletion loops

In the process of DNA replication, the primer and the template strand can separate. If the strands re-anneal incorrectly, insertions or deletions can occur. These are recognized and repaired by the MMR system (adapted from Levinson and Gutman, 1987).

3.3.1 DNA mismatch repair mechanism in *Escherichia coli*

The mechanism of DNA mismatch recognition and repair was first elucidated in studies with *Escherichia coli* (*E. coli*). Genetic screens, in which strains with "mutator" phenotypes were suggested to be MMR-deficient, helped identify key players of DNA MMR. Finally, the factors encoded by the identified genes were purified and the MMR system could be reconstituted *in vitro* (24). The first step of the MMR mechanism consists of recognition of the replication error, such as a base-base mismatch or IDL by the MutS homodimer, followed by binding of a MutL homodimer. The complex formation is ATP-dependent and allows its translocation on the DNA towards the closest hemi-methylated GATC site bound by the endonuclease MutH, which becomes activated. The hemi-methylated GATC site plays an important role in strand discrimination in *E. coli*. In order to maintain genome instability and prevent mutations, the newly synthesized DNA strand that contains the replication error has to be degraded. The enzyme responsible for methylation on the adenine, deoxyadenine methylase (Dam), is

slower than the replication machinery, resulting in transiently unmethylated GATC sites in the newly synthesized strand. Activated MutH incises the unmethylated GATC and UvrD helicase unwinds the nicked error-containing strand. Now, one of several exonucleases (ExoI, RecJ, ExoVII or ExoX) can degrade the strand beyond the mismatch, which could be up to 1 kb. The excision can be either 5'-3' or 3'-5' depending on the localization of the mismatch. The resulting gap is then resynthesized by DNA polymerase III and the remaining nick is sealed by DNA ligase (13,23).

It turned out that the main factors of MMR are highly conserved except for MutH. Only Gram-negative bacteria express MutH and no functional homologue was found in other organisms. Interestingly, MutH negative *E. coli* strains can repair mismatches under the condition that the DNA substrate contains a nick close to the mispair. This observation led to the conclusion that DNA nicks in general might provide the discrimination signal and the entry point for MMR. The key question is: Where do the nicks come from? An elegant solution is given by the DNA replication process itself: During replication, the lagging strand contains free termini at the 5' or 3' ends of Okazaki fragments and in the leading strand, at the 3' end (13,25).

3.3.2 DNA mismatch repair in humans

DNA mismatch repair is highly conserved in different organisms. Functional homologs of the bacterial MMR factors could be identified in yeast, humans and other species (23). Moreover, the mechanism of MMR is similar and studies from bacteria and yeast helped us understand mammalian MMR. In mammals, a mismatch or a small IDL are recognized by the heterodimer MutSα, which consists of MutS homolog 2 (MSH2) and MutS homolog 6 (MSH6). IDLs longer than 2 nucleotides are bound by MutSβ, a heterodimer of MSH2 and MutS homolog 3 (MSH3) (13). MutSα contains an ATPase domain (Walker ATP-binding motif) and biochemical studies with nuclear extracts and recombinant proteins showed that MMR is ATP-dependent. MutSα has strong affinity for mispairs in its ADP-bound state. Upon binding, ADP is replaced by ATP, which

brings about a rapid conformational change of MutSα and converts the complex into a sliding clamp on the DNA (23,25).

Similar to MMR in *E. coli*, MutSα recruits MutL homologs to the site of DNA damage. The most important MutL homolog is MutLα, which is again a heterodimer of MutL homolog 1 (MLH1) and PMS1 homolog 2 (PMS2). The name was derived from its homolog in yeast PMS1, which was discovered before the human protein. MLH1 also forms heterodimers with PMS1 and MLH3, which are called MutLβ and MutLγ, respectively. While the function of MutLβ remains enigmatic, MutLγ plays a role in meiotic recombination and can serve as a backup for MutLα (13,25).

In mammalian MMR, MutSα together with MutLα are believed to slide along the DNA until they encounter replication factor C (RFC) and proliferating cell nuclear antigen (PCNA) bound to a nick, which serves as the strand discrimination signal. RFC is removed and EXO1 is loaded to degrade the strand with the replication error. Interestingly, it turned out that degradation in both directions required only Exonuclease-1 (EXO1). Because EXO1 is a 5'-3' nuclease, it remained unclear how 3'-5' directed MMR functions in mammalian cells. Prominent explanations included the 3'-5' proofreading activities of DNA polymerases δ (Polδ) and ε (Polε), the 3'-5' exonuclease activity of MRE11 or the involvement of RFC (13). In 2006, Kadyrov and colleagues solved the problem when they discovered an endonucleolytic activity in the MutLα complex. It was localized at a $DQHA(X)_2E(X)_2$ motif in PMS2. With this observation, the model changed in a way that MutLα, after encountering the strand discrimination signal protected by PCNA and RFC, becomes activated and slides along the DNA to set additional nicks, which can be used as entry points for EXO1. In the case of 3'-5' directed MMR, nicks have to be set beyond the mispair in order to include the error in the degradation (26). The resulting single-strand gap is protected by replication protein A (RPA). As soon as the mismatch is removed, MutSα and MutLα dissociate from the DNA and EXO1 function is inhibited. Polδ is loaded at the 3' terminus by interaction with PCNA and resynthesizes DNA to fill the gap. Remaining nicks in the duplex are sealed by DNA ligase I (13,23).

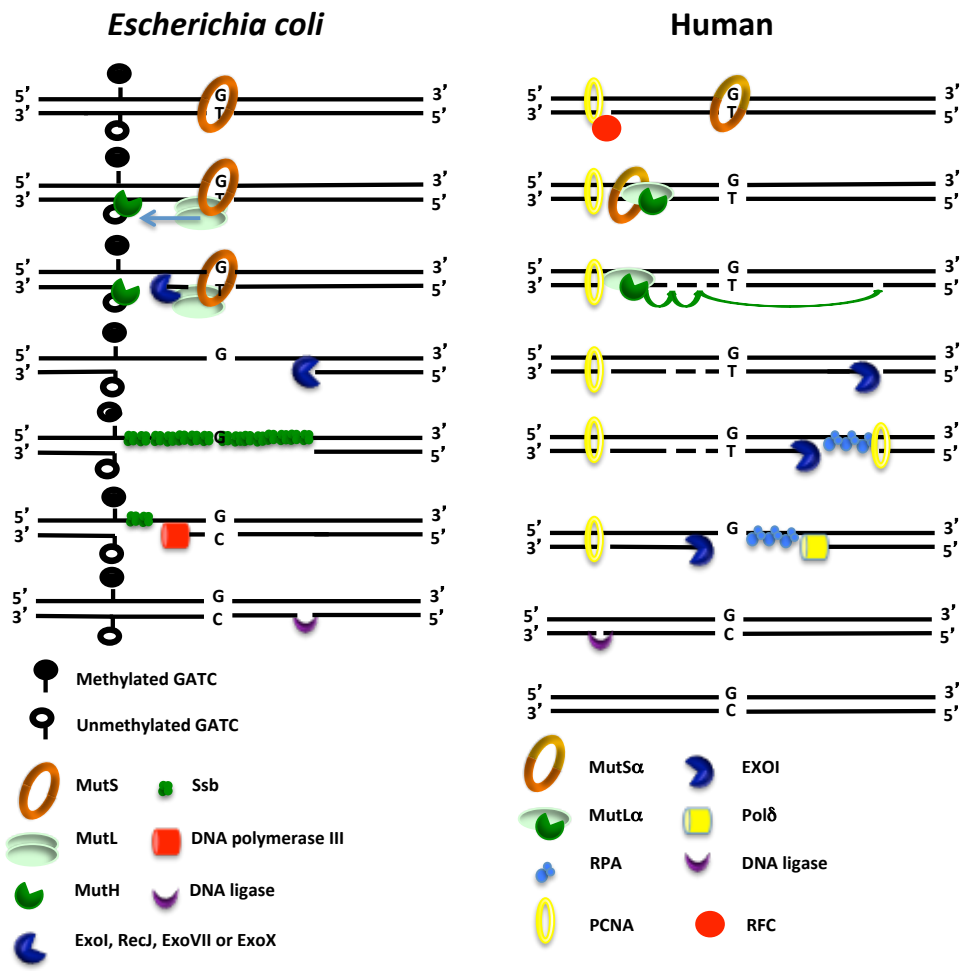


Figure 4: Mismatch repair in *Escherichia coli* and Humans

In *E. coli*, the replication error is recognized by MutS which than recruits MutL. Together, they slide along the DNA to the closest GATC site to activate MutH, which incises the unmethylated strand. Depending on the location of the mispair relative to the GATC site exonucleases ExoI or RecJ degrade the strand 5'-3' and ExoVII or ExoX degrade the strand 3'-5' beyond the mismatch, respectively. Ssb binds to the single stranded gap and DNA polymerase III is recruited to resynthesize the DNA duplex. The remaining nick is sealed by DNA ligase. In humans, the DNA mismatch is bound by MutSα, which recruits MutLα. They slide along the DNA until they encounter a nick. The MutLα endonuclease becomes activated by PCNA and RFC and starts to set additional nicks also beyond the mismatch. EXO1 is loaded and degrades the nicked strand in a 5'-3' direction. The single stranded DNA is protected by RPA until Polδ, recruited by PCNA, fills the gap. DNA ligase I seals the remaining nick in the DNA (modified from Jiricny, 2006).

As mentioned above, hemi-methylated DNA does not provide a discrimination signal in mammals due to the lack of a MutH homolog and therefore the explanation with nicks provided by Okazaki fragments or the 3' terminus in the leading strand is accepted also for MMR in humans (25). Over the course of the last few years, however, additional mechanisms that provide nicks for strand discrimination have been discovered. In 2013, two studies showed that ribonucleotides falsely incorporated during DNA synthesis are removed by RNase-H2 and the remaining nick can be used as strand discrimination signal for MMR (27,28). Also, 7,8-dihydro-8-oxo-guanine (8-oxoG), a result of oxidative stress, when incorporated into the newly synthesized DNA strand, can provide a signal for MMR when removed by factors of base excision repair (BER) that generate a nick in the DNA (29).

3.3.3 Regulation of the mismatch repair system *in vivo*

Most of the studies on the MMR mechanism were conducted in cell-free systems with plasmids as substrates for MMR. This helped identify the minimal requirements for MMR *in vitro*. In the context of chromatin, however, additional factors for efficient MMR might be required. Following this line, Li and colleagues identified an MMR-related interaction of MutSα with the histone mark H3K36me3. According to their model, MutSα is recruited onto chromatin through interaction of H3K36me3 with the PWWP domain of MSH6 (30). Recent data suggest that minichromosome maintenance complex component 9 (MCM9) plays a role in MMR. MCM9 might help unwind the DNA with its helicase domain for *in vivo* MMR (31). Other studies addressed the question of how MMR is timed in the context of chromatin assembly. MMR needs nucleosome-free DNA to repair replication errors. Recent data have shown that MMR factors travel with the replication folk (32) and therefore the replication machinery induces chromatin disassembly, which is needed for MMR. Reassembly of nucleosomes induced by chromatin assembly factor 1 (CAF-1) is rapid and leaves only a 250 bp long gap of naked DNA, which might not be enough for MMR. Data from Schöpf et al. indicate that MMR delays nucleosome assembly by interaction of MutSα with CAF-1 (33). In addition to timing and recruiting MMR factors, the

protein levels need to be tightly regulated. One mechanism is deacetylation and ubiquitination of MSH2 by histone deacetylase 6 (HDAC6), which leads to degradation of MSH2 (34). Moreover, MSH2 levels are regulated at the mRNA level by RNA interference (RNAi). MicroRNA-21 has been shown to be a potent regulator of MSH2 mRNA levels and microRNA-155 has been found to negatively regulate MSH2, MSH6 and MLH1 levels (35,36).

3.4 The role of mismatch repair in DNA damage repair and response

As the name suggests, the main function of the mismatch repair system is to take care of DNA mismatches that occur during replication including mispairs and IDLs. However, MMR factors have been shown to act in several DNA repair pathways and in DNA damage signaling. MMR was shown to play an important role in the repair of alkylation damage, oxidative damage and crosslinks within the DNA among others (37,38). In addition to active involvement in the repair of DNA lesions, MMR factors mediate DNA damage signaling by direct interaction with cell cycle checkpoint kinases such as ATR, Chk1 and Ck2 (37,39).

3.4.1 MMR-dependent repair of alkylation DNA damage

Alkylation DNA damage can be induced by several components, which are also among the oldest anticancer drugs. The most common compounds are monofunctional methylating agents such as N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), N-methyl-N-nitrosourea (MNU) or temozolomide (TMZ), bifunctional alkylating agents including nitrogen mustards chlorambucil and cyclophospamide, and chloroethylating agents such as nimustine or carmustine. Unimolecular nucleophilic substitution (S_N1) type methylating agents such as MNNG or MNU cause the formation of adducts at the N- and O-atoms of DNA bases. More than 80% of methylated bases consist of N-methylation adducts and can have long persistence in the DNA. For instance, N^7-methylguanine has a half-life of more than 80 h. O^6-methylguanine (^{Me}G) accounts only for about 0.3% to 8% of methyl adducts in the DNA, but they are very stable and persist in the DNA for a long time in the absence of O^6-

methylguanine-DNA methyltransferase (MGMT). Other adducts including N^3-methyladenine, N^3-methylcytosine and O^4-methylthymine are generated much less frequently. Different DNA repair mechanisms play a role in the removal of alkylation damage. N-alkylated bases are either repaired by BER or human AlkB homologues (hABH) (40). O6-^{Me}G is mainly repaired by MGMT. In a one-step repair process, the methyl group is transferred from the guanine in the DNA to a cysteine in the catalytic pocket of MGMT (41).

O6-^{Me}G lesions that are not repaired by MGMT, for instance in MGMT-deficient cells, are recognized as mispairs by MutSα (42). The modification takes place before replication and therefore O6-^{Me}G is found in the template strand. When MMR is activated, it will ultimately degrade the nascent strand with the C or T and the O6-^{Me}G remains untouched. The conclusive regeneration of the DNA strand results in O6-^{Me}G-C or O6-^{Me}G-T mismatches, which repeatedly reactivate MMR and resynthesis of the nascent strand, a phenomenon called "futile cycling". At some point, MMR might abandon the lesion but leaves behind single-strand gaps opposite O6-^{Me}G, which could cause double-strand breaks (DSB) in the next round of replication when the replication fork approaches. The formation of double-strand breaks leads to the activation of ATR and ATM checkpoint systems followed by cell cycle arrest in G_2/M (43,44). In addition, persistent repair of the O6-^{Me}G lesion might directly activate ATR (45). In contrast, cells deficient in MMR are not able to recognize the O6-^{Me}G mispairs and therefore do not undergo "futile cycling" and consequent ATR and ATM activation. In fact, MMR-deficient cells are up to 100 times more resistant to MNNG treatment than MMR-proficient cells (13).

3.4.2 MMR in intra- and interstrand crosslink repair

MMR-deficient cells are slightly resistant to *cis*-diammine-dichloro-platinum (II) (CDDP or known as cisplatin) (46,47). Cisplatin mainly causes 1,2-intrastrand crosslinks between two adjacent purines. Although MutSα has been shown to bind to 1,2-G-G adducts (37), the role MMR plays in the repair of these lesion remains largely unknown. Cisplatin-induced intrastrand crosslinks are mostly

repaired by nucleotide excision repair (NER) and cause a rapid S-phase cell cycle arrest. Some data suggest that MMR is activated by mismatches occurring during the repair of 1,2-G-G adducts and might therefore be involved in the induction of the cell cycle arrest (39,48).

For interstrand crosslinks (ICL), the picture is not quite clear. Studies using MMR-deficient cells revealed higher sensitivity against ICL inducing mitomycin C (MMC) compared to their matched MMR-corrected cells (49,50). In contrast, studies conducted with RNAi revealed a different situation. Cells with MLH1 knockdown behaved as wt or were slightly more resistant to MMC treatment (51,52). In some cases, MSH2 knockdown with RNAi led, however, to increased sensitivity (51). These differences suggest that canonical MMR is not essential for ICL repair, but that MMR factors might individually contribute to the repair.

3.4.3 ICL repair mechanism and contribution of MMR factors

ICLs are induced by agents such as MMC or nitrogen mustards, but also occur during normal cell metabolism. Defects in ICL repair result in *Fanconi anemia* (*FA*), which is characterized by an early onset of cancer, mainly acute leukemia, and bone marrow failure before the age of 45. In addition, most patients have congenital defects such as short stature, abnormal skin and developmental defects. Genetic studies revealed 19 FA or FA-like genes such as FANCA, FANCD2 or FANCI. In addition, factors from other DNA repair pathways were found such as the DSB repair proteins BRCA1 (FANCS) and BRCA2 (FANCD1) or the NER endonuclease DNA repair endonuclease XPF (ERCC4) (53).

The mechanism of ICL repair is intensely discussed and depends on the recognition pattern. Nevertheless, for all situations, repair of ICL lesions is a two-step process after recognition: Firstly, the ICL needs to be unhooked in one strand and secondly, the remaining adduct needs to be removed from the other strand. In replicating cells, most ICLs are recognized by approaching replication forks, which stall at the lesion. The FA core complex is recruited to the lesion and FANCD2 and FANCI are modified by ubiquitinylation and phosphorylation (53).

One model suggests that the leading strand is incised by the MUS81 structure-specific endonuclease subunit (Mus81) - essential meiotic structure-specific endonuclease 1 (EME1) and XPF - Excision repair cross-complementation group 1 (ERCC1), which generates a DSB at the replication fork. The gap is filled by translesion synthesis (TLS) polymerases such as Polζ together with REV1, Polκ, Polι or Polη**.** NER is activated to remove the single adduct and to fill the gap in the parental strand. With the help of homologues recombination (HR), the stalled fork is reconstructed and DNA replication moves on. In a situation where two forks converge, the first incision can occur on either strand. Again, the gap is filled by TLS polymerases. NER takes care of the remaining adduct and, with the help of HR, the replication fork is renewed (54,55). ICLs can also be recognized in the absence of DNA replication, either in the DNA duplex by NER with Xeroderma pigmentosum, complementation group C (XPC) as the main recognition factor or during transcription by a stalling of RNA polymerase. Xermoderma pigmentosum, complementation group G (XPG-1) incises 3' of the adduct and ERCC1 - XPF 5' of the adduct to unhook the ICL. TLS polymerases are recruited to fill the gapped structure. In the second repair step, the mono-adduct is recognized by a complex of damage-specific DNA binding protein 1 and 2 (DDB1 and 2) and NER is recruited to remove the adduct and resynthesize the DNA (56).

The role of MMR factors in the repair process is still under discussion. MutSα has been associated with the recognition of ICL lesions and DDR signaling, while another model predicts that the endonucleolytic function of MutLα can assist in the unhooking of ICLs at stalled replication forks. Alternatively, MLH1 might play a role in the recruitment of repair factors (54,56,57). MLH1 was shown to directly interact with FANCJ and FANCD2/FANCI-associated nuclease 1 (FAN1) (58,59). FANCJ is believed to be the preferred helicase to unwind the DNA after unhooking at stalled replication forks before DNA synthesis by TLS polymerases and FAN1 was shown to process ICL-like structures (55,60-62). Moreover, FANCJ- and also FAN1-deficient cells are hypersensitive to MMC treatment (63-65).

3.4.4 MMR in oxidative DNA damage and signaling

MMR repair factors MSH2 and MSH6 were shown to play an important role in the removal of DNA lesions caused by oxidative stress. Although the primary pathway of oxidative damage repair is BER, deficiency of MSH2 leads to increased levels of 8-oxoG in the DNA. An explanation for this observation was given by the Kannouche lab when they showed that, in response to oxidative stress, cells monoubiquitinylate PCNA to recruit the low-fidelity Polη to repair DNA lesions, which is dependent on MutSα but not BER (38).

Altogether, these observations made in different settings of DNA damage hint to an essential role of MMR factors not only in the repair of replication errors, but also in other aspects of DNA repair. In addition to the active repair of DNA lesions, MMR proteins are involved in DNA damage signaling. Consistent binding of MSH2 to DNA lesions, even without repair, can trigger the activation of cell cycle checkpoints by activation of ATR, CHK1 and CHK2, which are among the main signaling kinases in DDR (39). This activation leads to cell cycle arrest to gain time for repair of the lesions or, if the damage is beyond a certain threshold, it leads to apoptosis and cell death. Overexpression of MLH1 or MSH2 is sufficient for the induction of apoptosis; an observation that underlines the role of MMR factors in damage signaling (66). The molecular mechanism by which MMR triggers apoptosis is still under discussion. Treatment of cells with MNNG leads to phosphorylation of tumor protein p53 (p53) at serine 15 in MMR proficient cells and is reduced in MMR-deficient cells (67). Also, UVB-induced phosphorylation of p53 and induction of apoptosis is largely MSH2-dependent (68). However, cells lacking p53 also die efficiently after MNNG treatment in MMR-proficient cell lines which indicates two independent pathways (69). MMR was also implicated in the induction of tumor protein p73, the functional homolog of p53, in cisplatin-induced apoptosis. The accumulation and stabilization of p73 was shown to depend on MLH1 and PMS2 (70,71).

In summary, MMR factors clearly play a role in DNA damage recognition, cell cycle checkpoint activation and cell death, which seem to be independent of the

processing of the lesions. Interestingly, subnormal levels of MutLα are sufficient for efficient MMR, but for checkpoint activation, full levels of MLH1 are required (72).

3.4.5 Non-canonical mismatch repair as a source of mutagenesis

The mismatch repair system evolved to maintain genomic integrity by increasing the fidelity of DNA replication by two to three orders of magnitude. Hence, the idea that MMR might be involved in mutagenesis appears quite counterintuitive. Still, there is evidence that, under certain circumstances such as upon alkylating or oxidative damage, MMR factors trigger the ubiquitinylation of PCNA, which ultimately leads to the recruitment of low fidelity (error-prone) polymerase Polη (38,73). Under conditions of DNA damage, PCNA ubiquitination and subsequent Polη recruitment helps the cell to bypass lesions and to survive massive DNA damage, but at the cost of increased mutation rates. Not only DNA damage, but also the processing of U-G mismatches can trigger the ubiquitinylation of PCNA. Uracils in the DNA can arise by spontaneous deamination of cytosine or by incorporation of uridines during DNA replication and are mainly repaired by BER. However, if the BER pathway is inhibited, MMR is activated by U-G mismatches and processes the lesion with recruitment of Polη. Because non-canonical MMR is mutagenic, the question of the biological function besides cell survival arises. An explanation is given by the idea that increased mutation rates in B cells could help to increase antibody diversity during B cell maturation (74). Recent work by Bregenhorn et al. further suggests that MMR plays a role in generating DSBs needed for class switch recombination during B cell maturation (75).

3.5 Mismatch repair deficiency in carcinogenesis

Effective MMR is needed to maintain genome stability by reduction of mutation rates. Indeed, MMR increases the fidelity of DNA replication by up to 1000-fold (23). Constant absence of MMR results in higher mutation rates that affect all regions of the genome. Microsatellite regions are particularly prone to mutagenesis by deletion through forward slippage during DNA replication (76). The mutations do not only occur in non-coding regions, but also hit coding regions. Besides silent mutations, missense mutations that change the amino acid, non-sense mutations that lead to truncated proteins or frame shift mutations occur. When certain oncogenes or tumor suppressor genes are affected, these cells can acquire a selective advantage leading to increased growth and proliferation rates. This starts an evolutionary process which can mark the origin of a tumor.

Development of CRC is one of the best-studied models of tumor progression. However, MMR-dependent CRC development shows significant differences from the general model of CRC development. According to the hypothesis of Vogelstein and Kinzler, general CRC starts with mutations in *APC* or *β-catenin* (*CTNNB1*) genes followed by *kirsten rat sarcoma viral oncogene homolog* (*k-Ras*), *deleted in colorectal carcinoma* (*DCC*), *Mothers against decapentaplegic homolog 2* or *4* (*Smad2/Smad4*) and finally p53 at the transition from adenoma to carcinoma (Figure 5) (77). In contrast, CRCs that develop because of MMR deficiency establish a complex signaling network in between a series of activated and inactivated pathways including signal transduction (*TGFβRII*, *IGFIIR*, *PTEN)*, transcription regulation (*TCF4*), DNA repair (*MSH3*, *MSH6*, *MBD4*), apoptosis and inflammation (*BAX*, *CASP1*) (78-80). In humans, both alleles of an MMR gene in one cell have to be non-functional in order to inactivate MMR. In HNPCC, the first hit is a germline mutation (inherited from the parents) and the second hit a somatic mutation or LOH. In sporadic cancers, however, both hits are somatic.

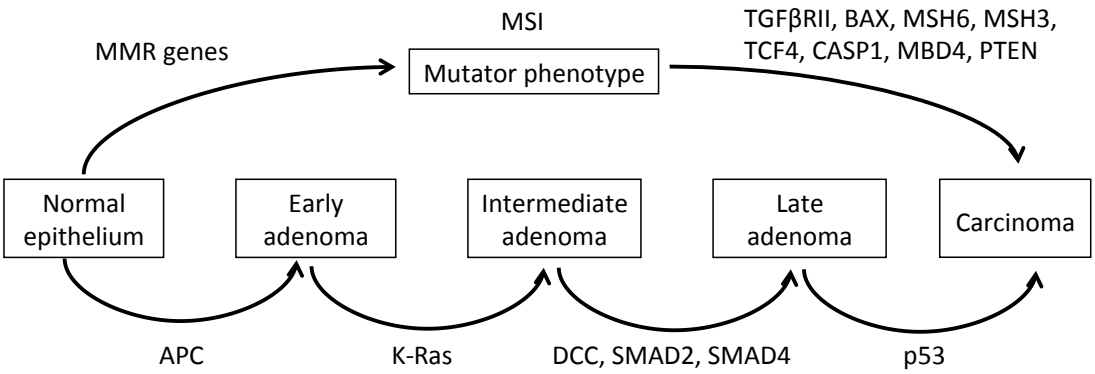


Figure 5: Development of colorectal cancers in MMR-proficient and -deficient cells
In MMR-proficient epithelial cells, the first mutation occurs in the *APC* gene, followed by mutations in *K-Ras*, which transforms normal cells and gives rise to early and later intermediate adenoma, respectively. Additional mutations in *DCC*, *SMAD2* or *SMAD4* and *p53* turn the lesion into late adenoma and finally into carcinoma. Normal epithelial cells that lose *MMR* genes develop a mutator phenotype and acquire mutations in many genes, but with mutation hotspots in *TGFβRII*, *BAX*, *MSH6, MSH3*, *TCF4*, *CASP1* and *PTEN*. Together, these mutations drive cancer development (adapted from Vogelstein and Kinzler, 1993 and Narayan and Roy, 2003 (81)).

3.5.1 Constitutive MMR deficiency

In HNPCC patients, only one allele of an MMR factor is mutated and a second hit is needed to inactivate MMR. Thus, only individual cells in particular tissues are affected and lead to tumor development. However, some individuals carry bi-allelic germline *MMR* gene mutations. Typically, such patients are offspring of consanguineous marriages in families with HNPCC history. Homozygous *MMR* gene mutations are associated with hematological malignancies and symptoms of neurofibromatosis type 1 (NF1) including neurofibromas, axillary freckles and café-au-lait spots. Obviously, timing of the inactivation has a large impact on the resulting phenotype and on the genotype as well (80,82). It appears that absence of MMR from the first stage of embryogenesis hits genes that are particularly fragile such as the NF1 gene or genes that are expressed in highly proliferating cells like hematopoietic cells. Only a few gastrointestinal cancers have been described in bi-allelic carriers. A possible explanation is given by the fact that all the CRCs were diagnosed after the age of nine. Most patients, however, die before the age of five and thus gastrointestinal malignancies do not have time to develop.

The phenotype of mice strains with MMR deficiency is consistent with the development of tumors in homozygous patients. The mice develop lymphoma malignancies after two to five months of age. Gastro-intestinal tumors are rare and are only observed in older mice (83). Mice heterozygous for *MMR* genes do not show any phenotypes. The difference between mice and humans might be explained by the short life span of mice, which is not long enough for the second hit to occur (84).

3.6 HNPCC-associated germline mutations

Genetic data from HNPCC patients and families have been collected over the past few years and are available on various databases including InSIGHT (16). For each HNPCC gene, the number of patients and variants as well as information on the variant such as MMR function, protein expression or protein localization is recorded. About 50% of the individuals have mutations in the *MLH1* gene (6637 of 13849) followed by mutations in *MSH2* with 37%. *MSH6* and *PMS2* are less frequently mutated (11% and 4%, respectively). In general, the mutations are spread over the whole sequence of the genes. A minor enrichment is found in exons 1 and 16 in MLH1 and exons 3 and 12 in MSH2 (15). The majority (about 80%) of the variants are unique in a way that they are found in only one HNPCC family. Missense mutations account for over 40% of all mutations in MMR genes (50% in *MLH1* and 34% in *MSH2*). Deletions, frame-shift and nonsense mutations occur together in about 30% of the patients (6/8%, 16/19% and 7/11% in *MLH1* and *MSH2*, respectively) (16).

3.6.1 Variety of MLH1 mutations in HNPCC families

The human *MLH1* gene is located on chromosome 3, spans a region of 55730 bp and consists of 19 exons. The transcribed mRNA has 2524 bp and its translation results in a protein with 756 amino acids and a molecular weight of 84.6 kDa (85,86). The MLH1 protein can be divided into two main functional domains: The amino terminal region contains the ATPase domain, which binds and hydrolyzes ATP, and the MutS homologs interaction domain; the carboxy terminal region consists of binding sites for PMS2, EXO1 and other MutL homologs (Figure 6). The interaction sites for PMS2 are particularly important, because dimerization with PMS2 results in the formation of MutLα, one of the main factors in MMR (87-89). The last 13 aa form the C-terminal homology (CTH) motif, which is highly conserved in eukaryotes. Crystal structures of yeast MLH1 revealed that the CTH motif is closely packed against PMS1 (PMS2 homolog in yeast) and that the last yeast MLH1 residue, Cys 769, participates in the PMS1 metal-binding site and hence contributes to the endonuclease function (90). Missense mutations that affect the function of MLH1 are found mainly in ATPase or PMS2 interaction

domains (11). Once an HNPCC family has been identified according to the Amsterdam criteria or Bethesda guidelines, genetic analysis is conducted to identify the underlying mutation (91). As a next step, functional testing is required to clarify the pathogenicity of the mutation. A wide range of methods has been used over the years to characterize patient *MMR* gene variants including testing of patient samples such as immunohistochemistry, genetic studies in yeast to identify mutator phenotypes, *in vitro* MMR assay and expression in human cell lines to analyze the expression pattern and location of the variant (11,92-94). For the majority of the variants, these techniques were suitable to explain the pathogenicity of the mutation. For example the MLH1 variants K84E and F80V show stable protein expression but are MMR-deficient in *in vitro* MMR assay (11). Because both mutations lie within the ATPase domain, it seems that they negatively affect the ATPase function of MLH1, though this has not been proven. In contrast, in the variants S247P and H329P, the mutations are outside of functional domains, but still cause strong MSI. When expressed in human 293T cells, strongly reduced protein levels could be observed. Mutations located in the PMS2 interaction domain were found to cause deficiency in binding to PMS2, which prevents the formation of the MutLα complex and consequently impairs MMR activity. There are, however, variants for which the pathogenicity has not been clarified yet. One of those is the mutant R659Q. Although the variant has been identified in 24 patients with early onset of CRC, the molecular pathology is still unclear (11,12,16,95).

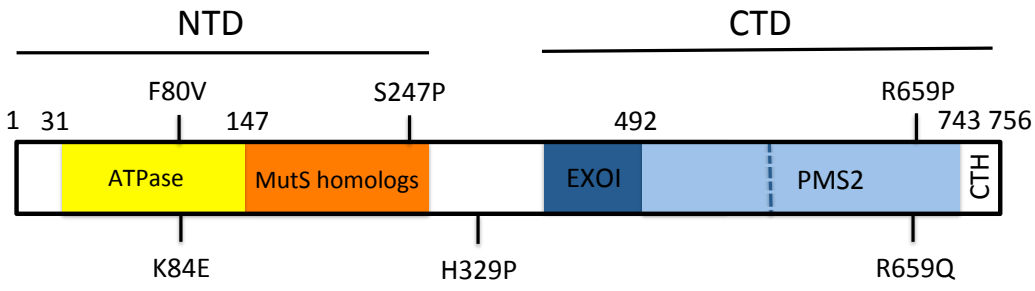


Figure 6: Structure of human MLH1

The human MLH1 protein is 756 amino acids long and contains two main domains: the N-terminal domain (NTD) that contains the ATPase domain and the interaction domain for MutS homologs as well as the C-terminal domain (CTD) which harbors the EXO1, PMS2 and other MutL homologs interaction domains and the CTH motif.

3.6.2 Clinical studies to identify *MMR* gene mutations predisposing for HNPCC

Once a mutation has been identified in an HNPCC patient and their family, it is of significant importance to gain information about its molecular pathogenicity. It is required to verify that the mutation is indeed the cause of HNPCC in the respective family, which provides essential information to screen other family members. Moreover, it helps to estimate the risk of each patient and might even help to develop specialized treatment options for each family, a process also known as personalized medicine. Before the functional characterization of genetic variants, segregation studies were conducted to distinguish pathogenic mutations from polymorphisms. Those studies try to segregate an amino acid change with the phenotype of the family, which indicates that a variant is a pathogenic mutation but does not prove it. Only in the case of mutations that bring about premature termination of translation by the creation of stop codons can the molecular mechanism be anticipated. However, functional testing might still be required. In HNPCC, patient numbers for each variant are very small and a lack of clinical samples makes it difficult to conduct reliable segregation studies (8,14).

The main molecular feature of HNPCC is MMR deficiency, which leads to MSI. Hence, tumor samples provide essential information. Specimens of colon cancers are checked for MSI and putative HNPCC patients are chosen for further studies, if they fulfill Bethesda guidelines developed by the InSiGHT study group: 1) The patient was diagnosed with CRC before the age of 50; 2) Presence of synchronous, metachronous colorectal or any other HNPCC-associated tumors regardless of age; 3) CRC with MSI-high status in a patient younger than 60; 4) CRC diagnosed at least in one first degree relative with HNPCC-related tumors with one of the cancers being diagnosed before the age of 50; 5) CRC diagnosed in at least two first- or second-degree relatives with HNPCC-related tumor independent of age (14). For MSI analysis, the National Cancer Institute microsatellite marker panel also known as Bethesda panel is widely used and contains five markers. According to the number of unstable markers, the MSI

phenotype is classified as high in the case of at least 2/5 or 40%, low in the case of 1/5 or 20% and stable, if none of the markers is unstable. MSI is observed in about 15-25% of sporadic colorectal and endometrial cancers and more than 90% of CRCs with MSI-high are sporadic tumors. In sporadic cancers, epigenetic silencing of *MLH1* by promoter methylation mostly causes MSI. In HNPCC, the affected gene influences the MSI status: Mutations in *MSH2*, *MLH1* or *PMS2* usually show high MSI, while mutations in *MLH3* and *MSH6* result in variable MSI status from stable to high (15).

Besides MSI analysis, tumor samples have been used for immunohistochemical (IHC) studies. IHC allows for the detection of a protein in a tissue with the help of antibodies and can provide information about the pathogenicity of a certain mutation. For instance, the absence of MMR factors in IHC hint towards a mutation that truncates the protein, affects the stability or other modes of gene silencing such as inactivating methylation. While IHC for MSH2 has been successfully established in many laboratories and proven to be reliable, IHC for MLH1 is more variable and it seems more difficult to interpret the results (96,97). To overcome these limitations, additional assays were developed to study the phenotype of mutations found in patients with a focus on those that do not cause premature termination.

3.6.3 Functional studies of HNPCC-related mutations

Several methods have been developed to study the phenotype of a given mutation in *MMR* genes and to provide information about the molecular pathogenicity. As the functional characterization is obvious for mutations that cause truncation of a protein, those methods were designed to study mutations that cause amino acid replacements and for which the phenotype is difficult to anticipate.

The first assays to study the phenotype of *MLH1* mutations were established in *Saccharomyces cerevisiae*. The first assay utilized the expression of human *MLH1* wt cDNA in MMR-proficient yeast strains, which resulted in a dominant mutator

effect judged by mutations in reporter genes for frameshift mutations, caused by the interaction of hMLH1 with the yeast MMR system. The introduction of selected MLH1 variants did not cause the mutator phenotype, which indicated a malfunctioning protein (93). The second yeast system utilized mutations in the yeast genome. Haploid as well as diploid heterozygous strains were analyzed. Interestingly, increased mutation rates in diploid strains were only observed when the wt allele was lost, which is consistent with observations in patients and indicates a dominant phenotype of the wt allele over the mutants. A clear advantage of the second system was the analysis of a yeast protein in a yeast system, which strongly improved the homology and the physiology of the system. Unfortunately, in this system only mutations that altered conserved residues between yeast and human could be studied (92). This drawback was overcome by a constructed human-yeast hybrid MLH1 protein introduced into yeast (98).

Another approach to study the influence of *MMR* gene mutations was based on glutathione-s-transferase (GST) pulldowns. GST fusion proteins with either MLH1 or MSH2 were created to test the complex formation of MutLα or MutSα, respectively (99,100). Although this assay is capable of detecting mutations that abrogate MutLα or MutSα formation, it fails to identify mutations that affect MMR function in another way, such as ATP metabolism or substrate interaction. Also, co-immunoprecipitation (Co-IP) methods were used to study the interactions of MMR factors MLH1, MSH2 and MSH6 (101-103). As it is laborious to create and analyze large numbers of MMR variants for GST pulldowns and Co-IPs, Kondo and colleagues developed a yeast two-hybrid system to study the interaction of many MLH1 variants with PMS2 and EXO1 (104). Interestingly, they found that also mutations in the N-terminal region of MLH1 abolished the interaction with PMS2 and EXO1, which is contradictory to previous data that localize the interaction domain to the C-terminus (87,100,105).

A very elegant method to study the function of MLH1 variants is provided by the *in vitro* MMR assay. Originally, nuclear extracts from cell lines were used to correct DNA heteroduplexes (106,107). Later, it was found that the extracts can

be complemented with either recombinant MMR proteins or whole cell extracts (108-110). The *in vitro* reconstituted system developed by the group of Paul Modrich provides another method to study the function of MMR factor mutants (24). To date, the *in vitro* MMR assay has proven to be the most sophisticated technique, as it gives information about the ability to perform MMR in a human system. However, it cannot provide information about the reason for MMR deficiency such as ATPase deficiency, interaction deficiency or protein localization errors.

Over the course of time, many different methods were developed to study the phenotype of *MMR* gene mutations with the *in vitro* MMR assay regarded as the most informative one when it comes to clinical usage. However, none of the systems alone is capable of providing comprehensive information about all aspects of a given mutation that leads to HNPCC. So far, only the combination of different methods was sufficient to classify the vast majority of *MLH1* variants as pathogenic or non-pathogenic. However, for many variants, e.g. R659Q, the classification is still uncertain and it requires more investigation or the development of new systems to study the phenotype (11,12,16,95).

3.7 The MLH1 interactome – methods and implications

Although the techniques described above helped to understand the molecular pathogenicity of many HNPCC families, the focus was mainly on the function of MLH1 in MMR. However, *MMR* genes have also been shown to play important roles in other aspects of DNA repair (13,54,59). The function or the mode of action of a given protein can often be anticipated if its interacting proteins are known. For MMR factors, several studies were conducted to identify new interaction partners and a wide range of methods has been applied including yeast and bacteria two-hybrid screens, pulldowns and Co-IPs combined with Western Blot or mass spectrometry (58,100,104,109,111). The results of those efforts associated *MMR* genes with other DNA repair pathways such as oxidative damage or ICL repair. But also interactions with proteins that are not primarily involved in DNA repair have been observed. For instance, an interaction of MLH1 with Actin has been observed and further studies suggest a role in cytoskeleton organization (111).

3.7.1 Two-hybrid systems

The two-hybrid system initially developed by Fields and Song in 1989 is based on two fusion proteins derived from the transcription factor GAL4 (112). Protein X is fused to the DNA-binding domain (BD) of GAL4 that binds to Gal upstream activating sequence (UAS_G) sequences, while protein Y is fused to the transcriptional activator domain (AD). Only if X and Y can form a stable complex is the transcriptional activator domain recruited to the DNA and transcription of the reporter gene downstream of UAS_G occurs. Although the method has been applied in other organisms such as *E. coli* (113), for protein-protein interaction screens mostly yeast is used. In principle, the protein of interest referred to as "bait" is fused to the BD, and a cDNA expression library encoding "prey" or "hit" proteins is fused to the AD. Both plasmids are introduced simultaneously and, upon interaction, the reporter gene is expressed. Most screens are based on cell survival. A mutant strain is used that is lacking an important enzyme for the biosynthesis of certain nutrients (amino acids or nucleic acids). If this mutant is grown on media without those nutrients, it fails to survive, unless the missing

factor is expressed because of the hybrid system. Surviving cells are analyzed and implicate an interaction of the bait protein with the prey protein (114). Table 1 shows MLH1 interaction partners that were identified using two-hybrid systems.

Protein	Role
BLM	Bait/Hit
EXO1	Bait/Hit
MBD4	Prey
MLH3	Bait/Hit
MSH3	Prey
MYC	Bait/Hit
ORC4	Bait/Hit
PMS1	Prey
PMS2	Bait/Hit

Table 1: MLH1 interacting proteins that were identified using two-hybrid system (87,111,115-120).

3.7.2 Pulldown assays and mass spectrometry

Genetic studies and the *in vitro* reconstituted system helped to understand the mechanism of MMR. It was anticipated that most factors identified in those studies would interact with each other. Indeed, when pulldowns and Co-IPs were performed, interaction of MutSα with MutL complexes could be verified (121). Also, the interaction of MutS as well as MutL complexes with EXO1 could be shown using the glutathione S-transferase (GST) pulldown system (105,122). In Co-IP assays, antibodies are used to capture specific proteins. Later on, the antibodies with the bound protein complexes are purified from cell extracts with Protein A- or G- coupled resins. In contrast, pulldown assays are based on tags fused to the protein of interest. GST-, Flag-, human influenza hemagglutinin (HA-) and histidine (His-) tags are among the most prominent ones. Instead of an antibody and Protein A or G coupled resins, molecules with high affinity to the tag coupled to resins are used. For instance, GST binds strongly to glutathione (GSH) and a matrix coupled with GSH is used for affinity purification of GST-tagged fusion proteins and its interaction partners. The purified complexes can be separated by sodium dodecyl sulfate (SDS) polyacrylamide gel

electrophoresis (PAGE) and specific proteins can be detected using Western Blot. Co-IPs and pulldown assays were used to show the interaction of MSH6 with the chromatin assembly factor CAF-1 and the histone mark H3K36me3 (30,33). The combination of Co-IPs or pulldowns and Western Blot, however, only allows for the identification of proteins one has been looking for, but it is not suited to identify unexpected or totally unknown interaction partners. This limitation has been overcome when pulldown assays were combined with mass spectrometric analysis.

Mass spectrometry provides a very powerful tool to identify unknown proteins. When pulldown assays and mass spectrometry are combined, referred to as affinity purification mass spectrometry (AP-MS), prey proteins can be identified. After purification of the complexes with the tagged bait protein, the mixture is digested with proteinases and the remaining peptides can be identified. Alignment of the peptides to databases provides information about the proteins that were bound to the bait protein and are considered as potential interactors. Cannavo et al. used TAP-tagged MLH1 and PMS2 to characterize the MutLα interactome (58). Table 2 provides an overview of proteins that were identified as MLH1 interactors using AP-MS.

Protein	Role
ATAD3A	Hit
ATAD3B	Hit
BLM	Bait
BRCA1	Hit
FANCJ	Bait
EXO1	Hit
FAN1	Bait/Hit
MMS19	Bait
MSH3	Hit
PMS1	Hit
PMS2	Bait/Hit
PP2A	Hit
P2BB	Hit
RUVBL1	Hit
RUVBL2	Hit

Table 2: MLH1 interacting proteins that were identified using AP-MS (58,59,123-127).

Although many interactors of MMR factors have been identified, the biological role of most of those interactions remains enigmatic and requires further investigation. Moreover, it is largely unknown in which pathways a certain interaction plays a critical role. Only a detailed study of the interactome of MMR factors in different aspects of DNA metabolism might help to decipher the many MMR interaction networks and to put them into functional context. In addition to the well-established methods, novel techniques might provide a promising approach. One possibility is the application of the BioID system developed by Roux and colleagues (128). The principle of BioID relies on a promiscuous BirA ligase, which generates reactive biotin that can bind to all proteins in close proximity. When the BirA ligase is fused to a protein of interest, its interaction partners become labeled with biotin and can later be isolated with streptavidin-coated resins. The Ting lab developed another interesting method; a mutated APEX peroxidase that generates reactive biotin-phenol which can also bind to nearby proteins (129). Again, a fusion protein could be used to label interaction partners of a given protein.

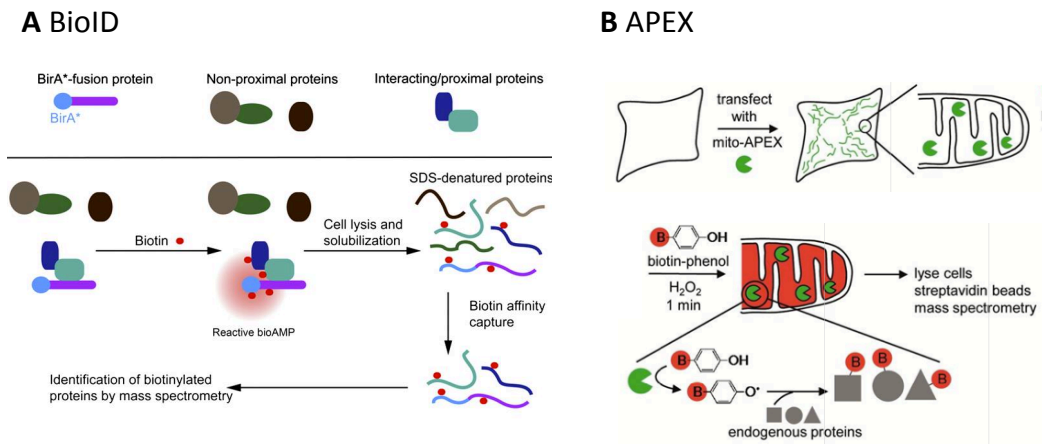


Figure 7: Principles of BioID and APEX

(**A**) BirA* tagged to bait protein generates reactive biotin that binds to proteins in close proximity. Bioinylated proteins are isolated with streptavidin resins and are identified by mass spectrometric analysis. (**B**) APEX protein with mitochondria localization signal generates reactive biotin-phenoxyl radicals that bind to proteins in close proximity. Labeled endogenous proteins are isolated and analyzed by MS (from Roux et al., 2012 and Rhee et al., 2013).

4 Aims of the Thesis

Many studies have been conducted to study the phenotype of gene variants found in patients with different hereditary diseases. Despite a lot of effort and the development of several methods, not all mutations could be classified as pathogenic or non-pathogenic. The first part of the present study aims to develop a universal tool suitable to study the phenotype of mutations found in patients with different diseases. The new system will be used to study the R659Q mutation of MLH1 found in HNPCC patients, which is of uncertain classification. The specific aims of the first part are as follows:

- Development of a general system to study the phenotype of germline mutations suitable for various diseases
- Application of the system to study the phenotype of the HNPCC-associated MLH1 R659Q mutation

In the second part, this thesis focuses on the development of methods to study the interaction of MMR factors in different aspects of DNA metabolism. The rationale for this investigation is given by the fact that most of what is known about MLH1 came from *in vitro* data and focused mainly on canonical MMR. With this study, we aim to identify new interaction partners of MLH1 *in vivo* and in other DNA repair pathways such as non-canonical MMR and ICL. Different approaches will be tested to identify stable interaction partners, transient interaction partners and factors that change upon induction of DNA damage. The specific aims of the second part are as follows:

- Generation of different systems to study protein interactions including BioID and APEX
- Interaction studies using mass spectrometry
- Follow-up of promising candidates

5 Material and Methods

5.1 Protein replacement with all-in-one plasmids

The all-in-one plasmid (pAIO) system provides an easy and reliable way to replace an endogenous protein with a protein variant such as a mutant or a tagged protein. The first pAIO system developed in the Jiricny group combines in one plasmid a cassette expressing a short-hairpin RNA (shRNA) to knock down the endogenous mRNA by RNAi and a second cassette expressing a cDNA, which allows for simultaneous expression of a gene variant (130). The shRNA is expressed from an H1 promoter that belongs to the RNA polymerase III class of promoters that drive the expression of short RNAs (e.g. rRNA and tRNA) and were shown to have ubiquitous robust activity in a wide range of cell lines. A CMV promoter, a widely used RNA polymerase II promoter, drives the expression of the cDNA. Importantly, both promoters contain tet repressor binding sites: one in the H1 promoter and two in the CMV promoter, respectively. These are of great value, as they allow for the induction of the system in cells that express the tet repressor such as T-REx-293. In addition, the plasmid contains a Puromycin cassette, which is used for the selection of stable cell lines. This first version of the system was used to replace the endogenous Polδ subunit p125 with variants that affect its function. Indeed, induction of the system led to higher mutation rates when the p125 subunit was replaced with either an error-prone variant or a proofreading-deficient variant and was even stronger in the double mutant. These experiments showed that the system works reliably in the replacement of endogenous proteins with variants.

Although the first version of the pAIO system worked efficiently, it expierenced some drawbacks, which could be overcome by improvements to the plasmid. The main features, the shRNA and cDNA expression cassette, were maintained and others were introduced into the plasmid (Figure 8). Firstly, attR sites for easier cloning using the Gateway technology (Invitrogen) taken from pEF5/FRT/V5-DEST plasmid (Invitrogen) were added. The Gateway technology is a powerful and efficient methodology for easy and reliable cloning. Once a cDNA is cloned

into an Entry Vector, it can be easily transferred to any destination vector by site-specific recombination with LR Clonase Mix. Moreover, we added an Flp Recombination Target (FRT) from pEF5/FRT/V5-DEST site for faster generation of stable cell lines. This is a great advantage over the first pAIO plasmids, as stable Flp-In 293 T-REx cell lines can be generated within three weeks. Moreover, due to the efficient recombination into a specific genomic localization, tedious picking of single cell clones becomes dispensable.

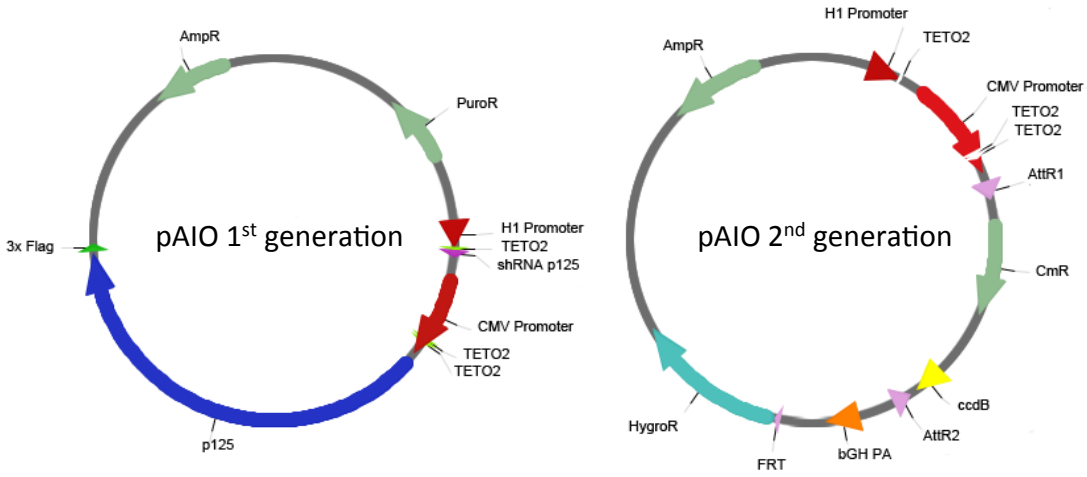


Figure 8: Comparison of 1st and 2nd generation pAIO systems

The main features of the pAIO system, the shRNA expression cassette with H1 promoter and tet repressor binding site (TETO2) as well as cDNA promoter expression cassette with CMV promoter and two TETO2 binding sites, were kept. The puromycin resistance gene (PuroR) for the selection of stable cell lines in the old pAIO plasmid was replaced by a hygromycin (HygrR) gene. In addition, the Gateway system with recombination sites (AttR1 and 2) and the ccdB toxin was inserted into the new plasmid. Moreover, the 2nd generation vector contains an FRT site for efficient integration into Flp-In cells.

5.2 Cloning methods for final pAIO plasmids

5.2.1 Insertion of shRNA into pAIO plasmids

In order to insert the shRNA into the pAIO plasmid, it was digested with restriction enzymes BglII and HindIII (NEB). As there is a stuffer sequence between the two restriction sites, the plasmid was run on an agarose gel to separate the fragments. The pAIO plasmid was purified with the QIAquick GEL Extraction Kit. The shRNA was ordered as DNA oligonucleotides and annealed as follows: 90°C 4 min, 70°C 10 min, 37°C 20 min, 24°C 20 min, 10 °C 10 min and 4°C forever. The shRNA oligonucleotide sequences were as follows: shRNA MLH1 cds forward 5' GATCCCCGGGTTACATATCCAATGCAAACTTCAAGAGAG TTTGCATTGGATATGTAACCCTTTTTA 3'; reverse complement 5' AGCTTAAAAAG GGTTACATATCCAATGCAAACTCTCTTGAAGTTTGCATTGGATATGTAACCCGGG 3', MLH1 3'UTR forward 5' GATCCCCGGGTGTTCTTCTTTCTCTGTATTCTCGAGAATA CAGAGAAAGAAGAACACTTTTTA 3'; reverse complement 5' AGCTTAAAAAGTG TTCTTCTTTCTCTGTATTCTCGAGAATACAGAGAAAGAAGAACACCCGGG 3'.

Importantly, the oligonucleotides had to have overhangs complementary to the sticky ends created by the restriction enzymes. The DNA oligos were ligated into the plasmid using T4 DNA ligase (NEB). The ligated plasmids were transformed into *E. coli* DH10b cells by electroporation and selected on agarose plates with ampicillin. Candidate clones were grown in 5 ml cultures and plasmids were extracted using NucleoSpin Plasmid kit. Positive clones were identified with DNA restriction using BglII and verified by sequencing (Microsynth).

5.2.2 Gateway cloning to insert GFP-MLH1 into the pAIO plasmids

The Gateway cloning system was used to integrate GFP-MLH1 into the pAIO plasmid. The entry vector containing GFP-MLH1 was mixed with the pAIO plasmid and LR Clonase Mix II (Invitrogen) was added for 1 h. Proteinase A was added for 10 min and the plasmids were transformed into *E. coli* cells and processed as described above.

5.2.3 Site-directed mutagenesis for shRNA-resistant cDNA and mutations

To mutate a specific site, primers complementary to the region close to the site to be mutated but containing the mutation are needed. After binding of the primers, Phusion DNA polymerase (NEB) synthesizes a new plasmid using the old one as a template. Several denaturation and amplification cycles enrich the mutated plasmid in the reaction. The conditions were the following: 95°C 2 min; 25x (95°C 1 min; 55°C 1 min; 63°C 25 min); 68°C 20 min; 4°C forever. In order to reduce the amount of non-mutated plasmid, Dpn I enzyme was added after the reaction. Dpn I exclusively degrades methylated DNA. The parental non-mutated plasmids produced in *E. coli* cells are methylated and will thus be degraded by Dpn I, while the newly synthesized plasmid, not methylated, is unaffected. The new construct was transformed into bacteria and the mutation was checked by sequencing after miniprep preparation. The primers to generate shRNA-resistant MLH1 cDNA were as follows: forward 5' GCCTTCAA AATGAATGGTTACATATCGAACGCGAATTACTCAGTGAAGAAGTGCATCTTC 3'; reverse complement 5' GAAGATGC ACTTCTTCACTGAGTAATTCGCGT TCGATATGTAACCATTCATTTTGAAGGC 3'. The primers to generate MLH1 mutants were as follows: K84E forward 5' GTGAAAGGTTCACTACTAGTG AACTGCAGTCCTTTGAGGATTTAG 3'; 5' reverse complement 5' CTAAA TCCTCAAAGGACTGCAGTTCACTAGTAGTGAACCTTTCAC 3', R659Q forward 5' GACTGCCTATCTTCATTCTTCAACTAGCCACTGAGGTGAA TTG 3'; reverse complement 5' CAATTCACCTCAGTGGCTAGTTGAAGAATGAAG ATAGGCAGTC 3'.

5.2.4 Polymerase chain reaction (PCR)

PCR is one of the most used methods in molecular biology and was developed by Kary Mullis in 1983 (131). It is used to amplify specific DNA regions. In principle, the DNA is first denatured at 95°C followed by cooling to allow for the annealing of primers complementary to a certain sequence and finally the elongation step, in which a DNA polymerase extents the primer by adding nucleotides 3' of the primer complementary to the template. This process is repeated up to 40 times, producing millions of copies in an exponential manner.

In this work, PCR was used to generate fragments of the CMV promoter. The reverse complement primer was the same for all and is located 3' of the TETO2 sites. The forward primer was different to generate fragments of different sizes to generate medium, low and very low expression CMV promoters according to commercially available Flexi Vectors from Promega. Restriction sites for AflII and HindIII were added to the forward and the reverse complement primers, respectively, for later insertion into the pAIO plasmid. The primer sequences were the following: forward medium 5' TAAGCAAAGCTTCCAAAATGT CGTAACAACCCG 3', low 5' TAAGCAAAGCTTGACGCAAATGGGCGGTAGGC 3', very low 5' TAAGCAAAGCTTATGGGCGGTAGGCGTGTACG 3' and reverse complement 5' CTGGACTAGTGGATCCGAGC 3'. Phusion polymerase (NEB) was used for the reaction and it was run at 95°C 1 min; 34x(95°C 30 sec; 62.5 30 sec; 72°C 1 min); 72°C 10 min; 4°C forever.

5.2.5 DNA purification, restriction, ligation and transformation

After the PCR reaction, the PCR products were purified using the QIAquick PCR Purification Kit (Quiagen) followed by restriction of the PCR products and the pAIO plasmid with restriction enzymes HindIII and AflII (NEB). The PCR products were again purified and the pAIO plasmid was run on an agarose gel to separate the plasmid backbone from the long CMV promoter fragment. The plasmid DNA was purified from the agarose band using the QIAquick GEL Extraction Kit (Quiagen). The different CMV promoter fragments were ligated into the plasmids and the ligated products were processed as described above.

5.3 Cell culture methods

5.3.1 Cell lines

Flp-In 293 T-REx cell lines were obtained from Life Technologies. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM, GIBCO-BRL) supplemented with 5% tet-approved fetal bovine serum (FBS-superior, Biochem AG), 100 U/ml penicillin, 100 µg/ml streptomycin (GIBCO-BRL) at 37°C and 6% CO_2. For the selection of stably transfected cells 250 µg/ml hygromycin (Invivogen) and 10 µg/ml blasticidin (Invivogen) were added to the medium. For activation of the pAIO system, Dox was added at a final concentration of 100 µg/ml for at least 4 days if not otherwise stated.

5.3.2 Cell transfections

Flp-In 293 T-REx cells were transfected using calcium phosphate as described (72). To generate stable cell lines, about 100,000 cells were seeded and 200 ng of pAIO plasmid and 1.8 µg of pOG44 (Flp-In recombinase) were transfected per 6-well after 24 h. On the next day, the cells were transferred to 10 cm dishes and the selection was started 24 h later. For transient transfections with Flag-FAN1 and Flag-FANCJ plasmids, 2 million cells were seeded and 40 µg of DNA was transfected the next day.

For transient RNAi treatments, 200,000 cells were seeded per 6-well and immediately transfected with Lipofectamine RNAiMAX (Life Technologies) and siRNAs obtained from Microsynth at a final concentration of 1 nM. The sequences for siRNAs were as follows: siMLH1 3'UTR 5' GUGUUCUUCUUUCUCUGUAUUdTdT 3'; siLuc 5' CGUACGCGGAAUACUUCGA dTdT 3'; siBCCIP1 5' GUGUGAUUAAGC AAACGGAUGdTdT 3'; siBCCIP2 5' GCCAUGUGGGAAGUGCUACdTdT 3'

5.3.3 Viability assay

200,000 cells were seeded per 6-well and treated with siRNA and Dox. After 48 h, cells were transferred into 96-well plates (500 cells/well) and again treated

with Dox and siRNA. 24 h later 10 µM O^6-Benzylguanine and MNNG or MMC were added. After 96 h of incubation time, cell viability was measured using Cell Titer Blue (Promega).

5.3.4 Cell cycle analysis

After the treatment, the cells were harvested and washed twice with cold PBS. The cells were resuspended in 70% ethanol and kept at -20°C for fixation. On the day of FACS analysis, the cells were washed with PBS once and resuspended in PBS containing 100 µg/ml RNaseA and 20 µg/ml propidium iodide. After incubation at room temperature for 30 min, the samples were analyzed using CyAn ADP (Beckmann).

5.4 mRNA quantification by real-time PCR

The cells were harvested and the total RNA was isolated using RNeasy Mini Kit (Quiagen). 1 µg of total RNA was converted into cDNA using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). cDNA was amplified and measured using Platinum® SYBR® Green qPCR SuperMix UDG (Invitrogen) and a LightCycler platform (Roche). The primers used for amplification were the following: GFP CAGTGGAGAGGGTGAA-GGTG and CGGGAAAAGCATTGAACAC; GAPDH CCCCGGTTTCTATAAATTGAGC and CACCTTCCCCATGGTGTCT.

5.5 Protein biochemistry

5.5.1 Whole cell protein extraction

On the day of protein extraction, the cells were harvested and washed with PBS. The cells were lysed with RIPA buffer (50 mM Tris pH 7.6, 150 mM NaCl, 0.1% SDS, 1% Trition, 0.5% Sodium Deoxycholate, 5 mM EDTA, 1mM PMSF, 1 mM DTT, 1x Roche Complete protease inhibitor cocktail, 1mM Na_3VO_4, 1 mM NaF) for 10 min on ice and centrifuged at 20,000 x g at 4°C. The supernatant was kept as whole cell extracts.

5.5.2 Nuclear extract preparation

Nuclear extracts were prepared as described previously (72). Briefly, about $8x10^8$ Flp-In 293 T-REx cells were washed in isotonic buffer and lysed in hypotonic buffer with five strokes of A-type pestle. The samples were centrifuged at 3,000 x g for 10 min and the supernatant was discarded. The nuclei were resuspended in extraction buffer and NaCl was added to a final concentration of 155 mM. Proteins were extracted through mixing for 60 min at 4°C. After centrifugation at 14,500 x g for 20 min the extracts (supernatant) were taken and dialyzed in extraction buffer followed by another round of centrifugation.

5.5.3 SDS-PAGE and Western Blot

After protein quantification using Bio-Rad Protein Assay (BIO-RAD), the samples were boiled with loading buffer for 5 min, separated by SDS-PAGE and blotted on polyvinyllidene diflouride membranes for 2 h at 100 V. After blocking with 5% non-fat powdered milk in 0.1% TBST (0.1% Tween-20 in TBS) for 30 min, the membranes were incubated with primary antibodies in 5% non-fat powdered milk in 0.1% TBST overnight at 4°C. The following antibodies and dilutions were used: MLH1 (Abcam, 1:2000), Lamin B1 (Abcam, 1:3000), β-Tubulin (Abcam, 1:3000), PMS2 (Santa Cruz, 1:250), MSH2 (Calbiochem 1:1000), MSH6 (BD Biosciences 1:1000), FANCJ (Cell Signaling 1:500), FAN1 (kindly provided by John Rouse 1:1000), TFIIH (Santa Cruz 1:1000) and Streptavidin-HRP (Invitrogen). The blots were washed 3 times with TBST and incubated with the appropriate secondary horseradish peroxidase-conjugated antibodies (GE Healthcare) for 1 h at room temperature. After additional washing steps, the proteins were detected with ECL (Advansta) and Fusion Solo imaging system (Witec AG).

5.5.4 *In vitro* MMR assay

The MMR *in vitro* assays were performed as described elsewhere (72). Heteroduplex DNA containing a G/T mismatch and a single-strand nick 307 bps 3' of the mismatch in the T-containing strand was incubated with 100 µg NE for 40 min if not stated otherwise. Repair of the mismatch restores a SalI restriction site and digestion of the plasmid with SalI and DraI results in three fragments, which were analyzed by agarose gel electrophresis. The amount of repaired substrate was quantified by normalizing the upper repair band to the remaining non-repaired band.

5.5.5 Pulldown assay

After treatment with Dox or transfection, the cells were lysed in NP40 lysis buffer (10 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP40, 1mM PMSF, 1 mM DTT, 1x Roche Complete protease inhibitor cocktail, 1mM Na_3VO_4, 1 mM NaF) for 30 min and were mixed repeatedly. The lysate was clarified by

centrifugation at 14,000 rpm for 10 min at 4°C. The protein concentration was measured using Bio-Rad Protein Assay (BIO-RAD). For pulldowns, 1 mg of protein in 1,000 µl dilution buffer (10 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1mM PMSF) was incubated overnight with 20 µl slurry of GFP-Trap-A (Chromotek) or Anti Flag M2-Agarose (Sigma) for GFP-MLH1 or Flag-FANCJ and Flag-FAN1, respectively. The resins were washed twice in dilution buffer and once in dilution buffer supplemented with 500 mM NaCl. To elute the bound proteins, the resin was boiled at 95°C in 2x loading buffer for 6 min. Samples were analyzed by SDS-PAGE and Western Blot.

5.5.6 GFP-Trap pulldown and mass spectrometry

In order to analyze interaction partners of tagged MMR proteins with mass spectrometry, cells were lysed with NP-40 buffer as described above. For chromatin extracts, cells were treated as described in preparation of NE, but chromatin-bound proteins were extracted with 420 nM NaCl. Instead of boiling in loading buffer, 50 µl 0.2 M glycine pH 2.5 were added for 1 min. Samples were centrifuged and the supernatant with eluted proteins was neutralized with 5 µl 1 M Tris buffer pH 10.4. For mass spectrometric analyses, the samples were first incubated with TCEP at a final concentration of 5 nM in Tris pH 8 for 20 min at 37°C followed by the addition of Iodoacetamide to a final concetration of 10 nM for 20 min. Proteins were digested with 1 µg trypsin (Promega) ON. On the next day, the digested peptides were cleaned up with ZipTip (Millipore) and dried in speed vac. The samples were resuspended in 5 µl 3 % acetontrile and 0.1% formic acid for processing in liquid chromatography – mass spectrometry (LC-MS), using Orbitrap Velos Mass Spectrometer (ThermoFischer). Data files were processed by the Functional Genomics Center Zurich and analyzed using Mass Scott and Scafold software.

5.5.7 Proximity-dependent *in vivo* biotinylation with BioID

Cells expressing BioID tagged proteins were incubated with biotin at a final concentration of 50 µM. They were then lysed with RIPA buffer followed by

Western Blot with Streptavidin-HRP (Invitrogen) to check for biotinylated proteins.

5.5.8 Proximity-dependent *in vivo* biotinylation using APEX

Cells expressing APEX2-tagged proteins with the pAIO system were generated (132). After induction of the system for 72 h, the cells were treated with biotinphenol at a final concentration of 500 µM for 30 min and 1mM H_2O_2 for 1 min. The cells were washed with Quencher solution (10 mM sodium azide, 10 mM ascorbic acid, 10 mM and Trolox 5 mM). The cell pellets were frozen in liquid N_2 and thawed in PBS. The samples were centrifuged for 10 min and 3000 x g at 4°C. Supernatant was taken as a cytoplasmic extract and the nuclei were washed once in PBS. The nuclei were resuspended in 0.5 volumes of the pellet nuclear extraction buffer and 0.031 volumes of 5 M NaCl were added. Nuclear proteins were extracted by rolling for 60 min at 4°C. The samples were either processed for Western Blot or LC-MS.

5.5.9 Pulldown and LC-MS of biotin-labeled proteins

Cytoplasmatic extracts and nuclear extracts were incubated with High Capacity Streptavidin Agarose (Pierce) ON at 4°C with shaking. The resins were centrifuged at 500 x g and washed twice with wash buffer 1 (dH_2O with 2% SDS), once with wash buffer 2 [0.1% deoxycholate, 1% Triton X-100, 500mM NaCl, 1 mM EDTA, 50 mM Hepes (ph 7.5)], once with wash buffer 3 [250mM LiCl, 0.5% NP-40, 0.5% deoxycholate, 1mM EDTA, 10 mM Tris (ph8.1)], and twice with wash buffer 4 [50 mM Tris (ph 7.4) and 50 mM NaCl]. In order to elute the bound proteins, the resins were boiled in 2x loading dye at 95°C for 5 min. For LC-MS, the agarose resins were treated as described (133). Briefly, the resins were intensively washed with 10 ml 0.5% Triton X-100 in 50 mM amonium bicarbonate, followed by 10 ml 5M NaCl, followed by 10 ml 100 sodium carbonate and finally 10 ml ammonium bicarbonate. For washing, the resins were transferred into Mobicols to a Vac-ManLaboratory Vacuum Manifold (Promega). The beads were treated with TCEP and Iodacetamide and the proteins were digested for 4 h with LysC (Calbiochem) and subsequently

overnight with trypsin (Promega). The peptides were cleaned up with Ultra Micro SpinColumn (Harvard) and analyzed in an Exactive Plus Orbitrap Mass Spectrometer (Thermo Scientific).

5.5.10 Protein identification and data filtering

For the AP-MS experiments with GFP-MLH1 and MSH2-eGFP acquired spectra were searched with Mascot software against the human proteome (www.uniprot.org) (134). The search parameters were set to only fully tryptic peptides with up to two missed cleavages. Oxidation on methionine and phosphorylation on serine, threonine and tyrosine were set as variable peptide modifications. Carbamidomethylation on cysteine was set as fix modification. Error tolerances on the precursor and fragment ions were ±25 ppm and ±0.5 Da, respectively. The obtained peptide spectrum matches were statistically evaluated using Scaffold software (v4.05) (135). Normalized total spectra (NTS) higher than five was considered as an interactor and lower than five was considered as negative resulting in a list of potential GFP-MLH1 interactors. In a next step the protein list of GFP-MLH1 was matched against GFP only. Proteins with NTS higher than five in the control were discarded.

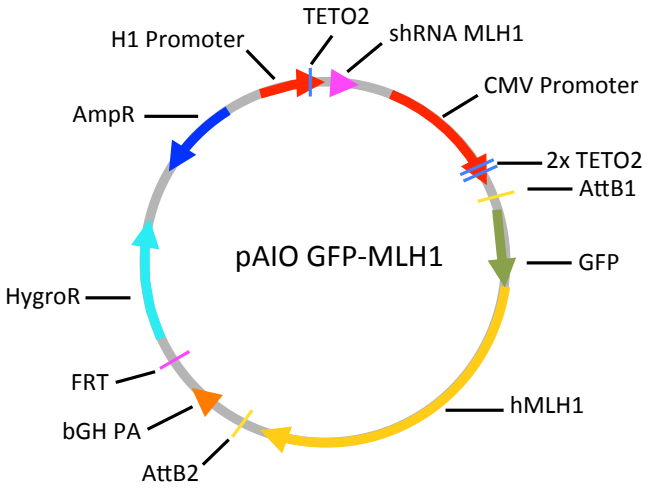
For APEX experiments a different approach was used. After data acquisition, the Prohits pipeline was applied (136). The obtained spectra were searched with search engines Comet and X!Tandem (137,138). The search parameters were set to only fully tryptic peptides and up to two missed cleavages. Carbamidomethylation on cysteine was set as a fix peptide modification. Oxidation on methionine and phosphorylation on serine, threonine and tyrosine were set as variable peptide modifications. The precursor mass tolerance was set to 20 ppm and the fragment mass error tolerance was set to 0.015 amu. Protein-level validation and protein inference of the individual search engines were performed with Peptide Prophet and combined in iProphet (139) using the Trans-Proteomic Pipeline (TPP, v4.7 POLAR VORTEX rev 1, Build 201510301006 (linux)). For further analysis only proteins with an iProphet probability of >0.95 and minimum two peptides were considered. In order to identify and estimate

high confident protein interactions, Saint-express software (Version:exp3.3) was used (140,141). The resulting list contained of potential APEX-MLH1 interactors with average probability (AvgP) and Bayesian false discovery rate (BFDR). AvgP of higher than 0.9 and BFDR lower than 0.02 were considered as statistically significant.

6 Results

6.1 Usage of second generation pAIO system to replace endogenous MLH1

The first aim of this study was to develop an easy and universal system that can be used to study the phenotype of mutations found in patients. In order to do so, the wild type protein needs to be inactivated or removed from the cell. The pAIO system makes use of RNAi through an shRNA to knock down the endogenous protein. We introduced two different shRNAs into the pAIO plasmid targeting either the cds or the 3'UTR of the endogenous MLH1 mRNA. We decided to use *MLH1* as a representative example for diseases that are caused by inherited mutations in one single gene to study the phenotype of certain mutations in the *MLH1* gene found in HNPCC patients. To differentiate the exogenous from the endogenous MLH1, we used GFP-MLH1 which has previously been reported to be functional (142). The GFP-MLH1 cDNA was introduced into the pAIO plasmid and stable cell lines were generated. Importantly, the GFP-MLH1 cDNA for the pAIO plasmid with the shRNA targeting the cds was made shRNA-resistant by mutagenesis while retaining the peptide sequence (Figure 9).



shRNA target against 3'UTR of MLH1 mRNA
GTG TTC TTC TTT CTC TGT ATT C

shRNA target against cds of MLH1 mRNA
GGT TAC ATA TCC AAT GCA AAC
shRNA-resistant target in MLH1
GGT TAC ATA TCG AAC GCG AAT
MLH1 polypeptide sequence
G Y I S N A N

Figure 9: Replacement of endogenous MLH1 with GFP-MLH1 using pAIO
Schematic representation of pAIO GFP-MLH1. Sequences of the H1 and CMV promoter, positions of the shRNA and the tet binding sites (TETO2), open reading frames for the GFP-MLH1 cDNA, antibiotic resistance markers (AmpR and HygroR), positions of the Gateway (AttB1 and 2) and FRT recombination sites and bovine growth hormone polyadenylation signal (bGH PA) are indicated. Sequences of the shRNAs used and the position of the silent mutation in the MLH1 cDNA, which does not alter the polypeptide sequence are shown below.

After transfection of Flp-In 293 T-Rex cells with the pAIO plasmids, stable cell lines were selected. Besides single cell clones, the bulk culture was also analyzed for MLH1 replacement. As shown in Figure 10 A, both shRNAs could knock down endogenous MLH1 efficiently and all single cell clones, as well as the bulk cultures (BC), expressed GFP-MLH1. The expression was very homogenous between the different cell lines, which led to the conclusion that BCs are suitable for the generation of stable cell lines and that picking of single cell clones was not necessary. Although the knock down was efficient, endogenous MLH1 could still be detected. It has been reported that only 10% of MLH1 protein is enough for

efficient MMR (72). Hence, the cells with shRNA targeting the cds were additionally treated with siRNA against the 3'UTR. The combination of both shRNA and siRNA significantly increased the knock down efficiency (Figure 10 B).

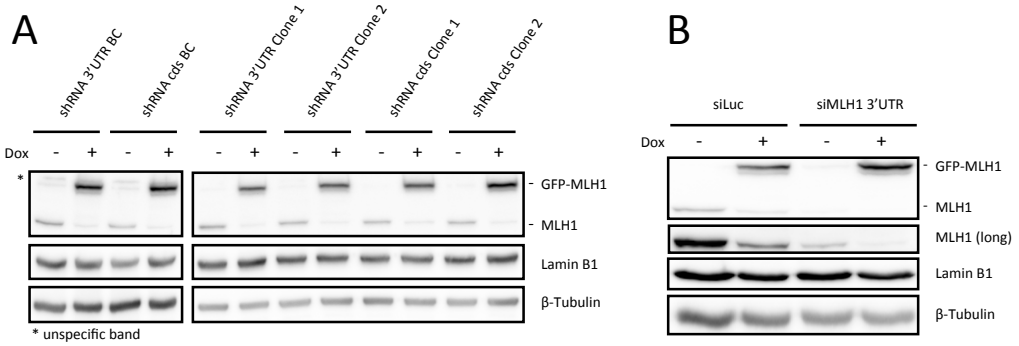


Figure 10: Overexpression of GFP-MLH1 and knock down of MLH1 in pAIO

(**A**) Western Blot of extracts of stable cell lines (bulk culture BC or single cell clones) expressing endogenous MLH1 (- Dox lanes) or the GFP-tagged variant (+ Dox lanes) with either the 3'UTR directed shRNA or cds directed shRNA upon treatment with Dox (+) for 96 h. (**B**) Combination of cds shRNA induced with Dox and treatment with siMLH1 against 3'UTR for 96 h substantially improved the knock down of endogenous MLH1.

In all cell lines analyzed so far, the expression of GFP-MLH1 was much stronger than that of the endogenous MLH1. Because overexpression of MLH1 was shown to impair MMR in yeast and to activate apoptosis in human cells (66,143), we thought about strategies to reduce the GFP-MLH1 levels down to those of endogenous MLH1. The analysis of an additional cell clone verified the high homogeneity between the clones and we concluded that we would not find clones with much less expression (Figure 11 A). Moreover, the reduction of the Dox concentration did not reduce the GFP-MLH1 levels (Figure11 B), but we found that a concentration which is 1000-fold (1 ng/ml) lower than that recommended by Invitrogen was sufficient to activate our system. When treated with 0.1 ng/ml Dox, the GFP-MLH1 levels were reduced, but the knock down of endogenous MLH1 was lost.

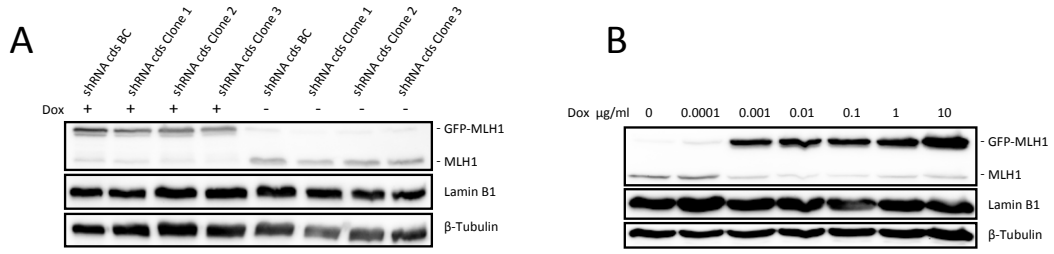


Figure 11: No reduction of GFP-MLH1 expression by low Dox concentrations

(**A**) Western Blot of bulk culture (BC) and 3 single cell clones with (+) and without (-) Dox (1 µg/µ) showed homogenous high expression of GFP-MLH1 among all cell lines (**B**) Titration of the Dox concentration did not reduce GFP-MLH1 expression in BC as judged by Western Blot.

A more elegant way to reduce GFP-MLH1 levels is the reduction of the mRNA expression at the level of transcription. This can be achieved by using promoters that are weaker than the CMV promoter. Instead of changing the whole promoter, we decided to shorten our CMV promoter according to Promega's CMV promoters with different expression levels. Using PCR, CMV promoter fragments of different lengths were generated and integrated into the pAIO plasmid (Figure 12 A). As shown in Figure 12 B, the GFP-MLH1 levels decreased with the length of the CMV promoter. Our data suggested that the medium CMV promoter gives rise to GFP-MLH1 comparable to endogenous MLH1.

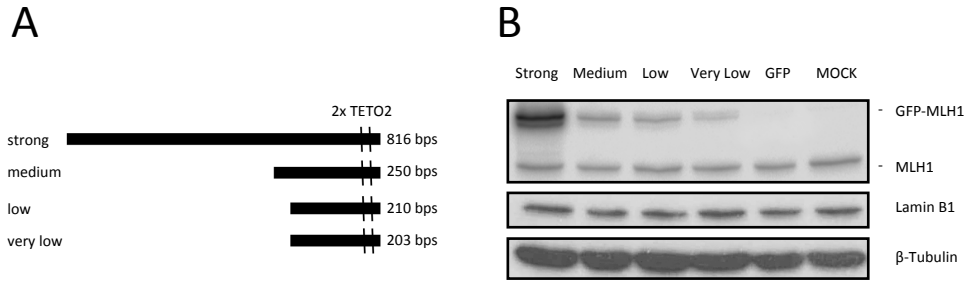
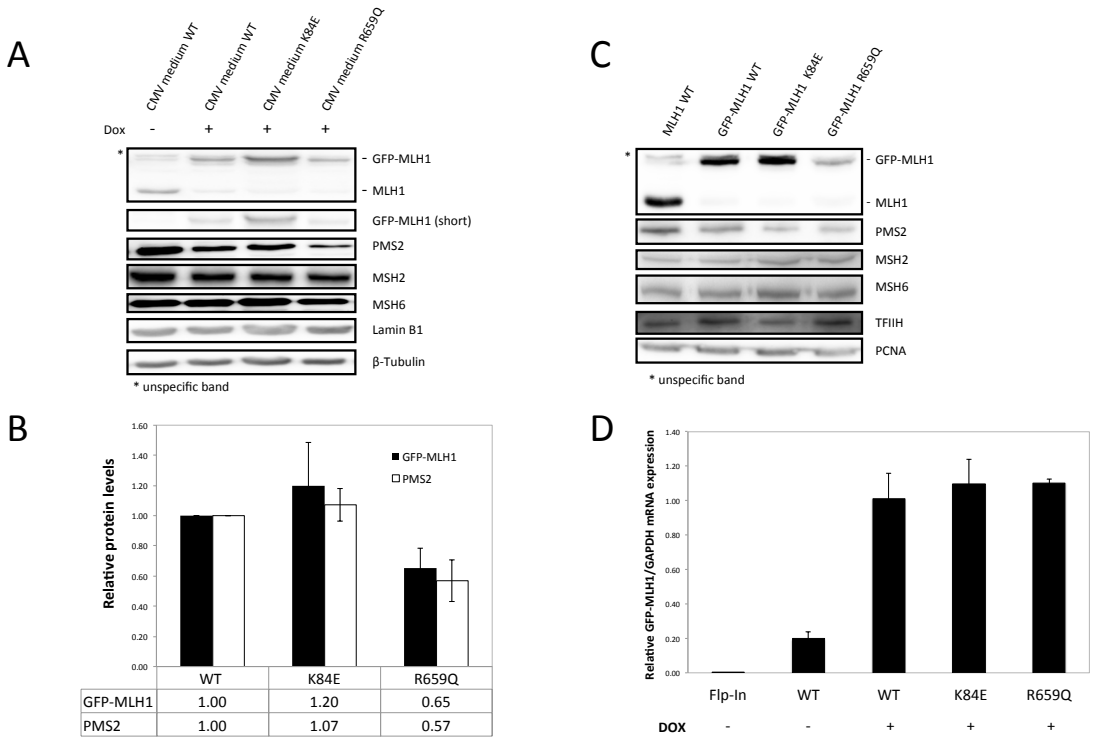


Figure 12: Reduction of GFP-MLH1 expression by promoter shortening

(**A**) With the help of PCR and standard cloning techniques, the CMV promoter in the pAIO plasmid was shortened according to the scheme. (**B**) CMV variants were transiently transfected into Flp-In 293 T-Rex cells and Western Blot analyses revealed reduced GFP-MLH1 levels in shorter CMV promoters. The proteins were extracted already 24 h after transfection to allow for comparison of GFP-MLH1 to endogenous MLH1 before the knock down by the shRNA became visible.

As the aim of the study was to analyze the phenotype of patient mutations, we then generated cell lines with medium CMV promoter and GFP-MLH1 K84E and R659Q. When comparing the protein levels, we found that the K84E mutant showed increased levels of GFP-MLH1 in whole cell extracts (WCE), but not in nuclear extracts (NE) (Figure 13 A-C). The protein levels of GFP-MLH1 (R659Q), however, were decreased in WCE and in NE. Moreover, the levels of PMS2 were reduced in WCE while other MMR factors were not. The mRNA levels were equal in wt and the mutants, which indicates that the R659Q mutation affects protein stabilization (Figure 13 D). We could not, however, exclude translational repression by RNAi.



	WT	K84E	R659Q
GFP-MLH1	1.00	1.20	0.65
PMS2	1.00	1.07	0.57

Figure 13: The MLH1 R659Q mutation leads to reduced GFP-MLH1 protein levels

(**A**) pAIO GFP-MLH1 wt, K84E and R659 with medium CMV were generated and GFP-MLH1 levels were measured by Western Blot 96 h after induction. (**B**) Quantification of GFP-MLH1 and PMS2 levels of three independent replicates and SD. GFP-MLH1 K84E levels were higher than wt, whereas the mutant R659Q revealed reduced GFP-MLH1 and PMS2 levels. (**C**) Nuclear extracts of the cell lines were prepared and the protein levels were analyzed. GFP-MLH1 WT and K84E showed similar levels to endogenous MLH1. The R659Q mutation also reduced the GFP-MLH1 levels in the nucleus. Note that the cells were additionally treated with siLuc (- Dox) or siMLH1 3'UTR (+ Dox) to maximize the knock down of endogenous MLH1 for WCE and NE. (**D**) mRNA levels were measured by qRT PCR of GFP-MLH1 in uninduced (- Dox) or induced (+ Dox) cell lines and no difference was observed. Non-transfected Flp-In 293 T-Rex cells were used as the negative control.

6.2 The K84E, but not the R659Q, mutant is MMR-deficient

In our next experiments, we focused on the MMR activity of the MLH1 mutants. Time course *in vitro* MMR assays revealed very low MMR activity of the K84E mutant (Figure 14 A-D). We also observed reduced activity of the R659Q mutant, but it was not significant over three independent experiments. To rule out that the MMR deficiency of the K84E mutant was caused by inactive nuclear extracts, purified Mutlα wt and PMS2 endonuclease-dead D699N mutant (DN) were added to the reaction (26). The wt complex could restore the MMR activity of the K84E extracts but the DN mutant failed to do so and thus implicated that the observed reduction of MMR activity is indeed caused by the K84E mutation, which is in line with observations by others (11,12,95). To verify the MMR activity of the R659Q mutant, we made use of 293T Lα cells and the R659Q analogue (72). Again, the R659Q mutant was active. Moreover, the pAIO GFP-MLH1 variants were tested for MNNG sensitivity (Figure 14 E). MNNG was shown to activate MMR and MMR-proficient cells are highly sensitive to the treatment. The R659Q mutant behaved exactly as wt GFP-MLH1, but both were slightly resistant compared to non-MLH1-replaced cells. In contrast, the K84E mutant was highly resistant which underlines its MMR deficiency. These results strongly suggest that the R659Q mutation, despite reduced protein levels, does not severely reduce MMR activity and we concluded that impaired MMR activity might not be the main cause for tumor development in patients with the R659Q mutation.

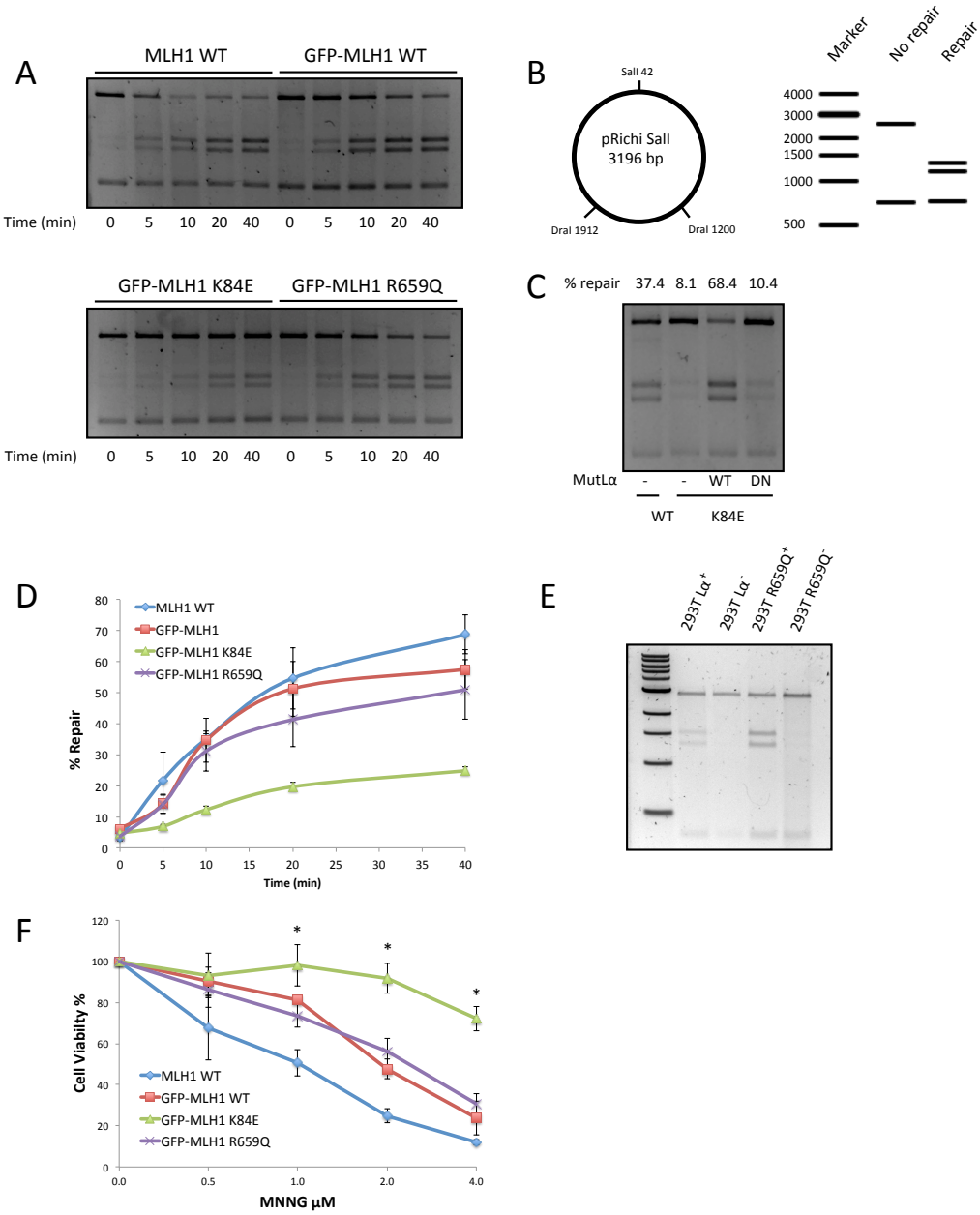


Figure 14: The R659Q mutation does not affect MMR activity

(**A**) MMR time course assay in NE from stable cell lines expressing endogenous MLH1 (MLH1 WT) or GFP-MLH1 variants (WT, K84E or R659Q) revealed no difference for the R659Q mutation. NE from the K84E mutant, however, was significantly less active. (**B**) Schematic explanation of *in vitro* MMR assay. (**C**) NE from K84E GFP-MLH1 are MMR active, when supplemented with WT MutLα, but not with PMS2 endonuclease-dead MutLα (DN). NE from GFP-MLH1 WT cells served as a positive control. (**D**) Quantification of three independent experiments and data are represented as mean ± SD. (**E**) NE from 293T Lα^{+}, 293T Lα^{-}, 293T R659Q^{+} or 293T R659Q^{-} cells were prepared and MMR assays performed. The results verified that the MLH1 R659Q mutant is MMR-proficient. (**F**) The K84E but not the R659Q mutation affects MMR activity *in vivo* as

judged by cell viability assay of stable 293 cell lines expressing endogenous MLH1 (MLH1 WT) or GFP-MLH1 variants (WT, K84E or R659Q) treated with 10 µM O^6-Benzylguanine for 1 h, followed by MNNG. Data are represented as mean of three independent experiments ± SD. Asterick indicates significant differences between GFP-MLH1 WT and K84E with a p-value below 0.05 according to t-test.

6.3 Long-term replacement with GFP-MLH1 K84E did not induce MSI

The previous experiments revealed that the K84E mutation, but not the R659Q mutation, impairs MMR *in vitro* as well as *in vivo* as judged by its resistance towards MNNG. As microsatellite instability (MSI) is a very prominent hallmark of MMR deficiency, we isolated the genomic DNA after four weeks of continuous induction of the pAIO system and checked for MSI. This was done in collaboration with Karl Heinimann from the Universitätsspital Basel. Genomic DNA from 293T cells, which lack MLH1, was used as positive control and showed strong shortening in micro satellite BAT40 compared to 293 MLH1 wt cells (Figure 15). The replacement of endogenous MLH1 with the GFP-MLH1 variants, however, did not affect the length of BAT40. Similar results were obtained for other microsatellites including BAT26, BAT25 and EWSR1 (data not shown).

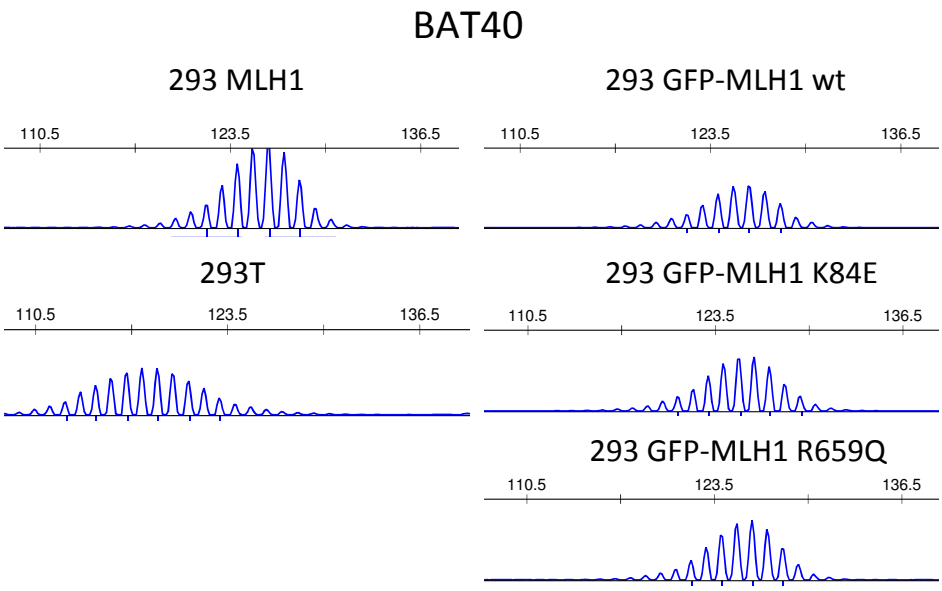


Figure 15: MSI analysis of GFP-MLH1 variants

The pAIO system was induced in all the variants for four weeks and genomic DNA was isolated and submitted for MSI analysis. The MSI BAT40 was chosen as a representative result for six markers in total.

6.4 The R659Q mutation impairs binding to FAN1 and FANCJ

MLH1 was shown to play an important role also in other aspects of DNA repair and its interaction with other repair factors is crucial for genome maintenance. Several studies suggest an involvement of MLH1 in ICL repair, which might be mediated by interaction of MLH1 with FAN1 and FANCJ (58,59,124). In order to analyze the binding of MLH1 to certain factors, we made use of the GFP-tag and performed a pulldown with GFP-Trap. As shown in Figure 16 A, all variants strongly interacted with PMS2, and GFP-MLH1 wt as well as the K84E mutant also bound FAN1 and FANCJ. The R659Q mutant, however, showed strongly reduced interaction with FAN1 and no visible binding to FANCJ. To verify these findings, FLAG-FANCJ and FLAG-FAN1 were expressed in the cell lines and pulldowns with Anti-FLAG M2 antibody were performed. Endogenous MLH1, GFP-MLH1 wt and K84E could be found in the pulldowns, but binding of GFP-MLH1 R659Q was significantly lower for both Flag-FANCJ and FLAG-FAN1 (Figure 16 B, C). When we did a similar experiment in 293T Lα and L R659Q cells

using a MLH1 antibody, we could verify that the binding of MLH1 R659Q to FANCJ and FAN1 was strongly impaired compared to MLH1 wt (Figure 16 D). These experiments made clear that the R659Q mutation substantially reduces the interaction of MLH1 and FANCJ or FAN1. Both of them have been shown to play an important role in the repair of ICL lesions in the DNA and one could imagine that ICL repair is altered in MLH1 R659Q cells, which could contribute to tumor development (59,62,65,124).

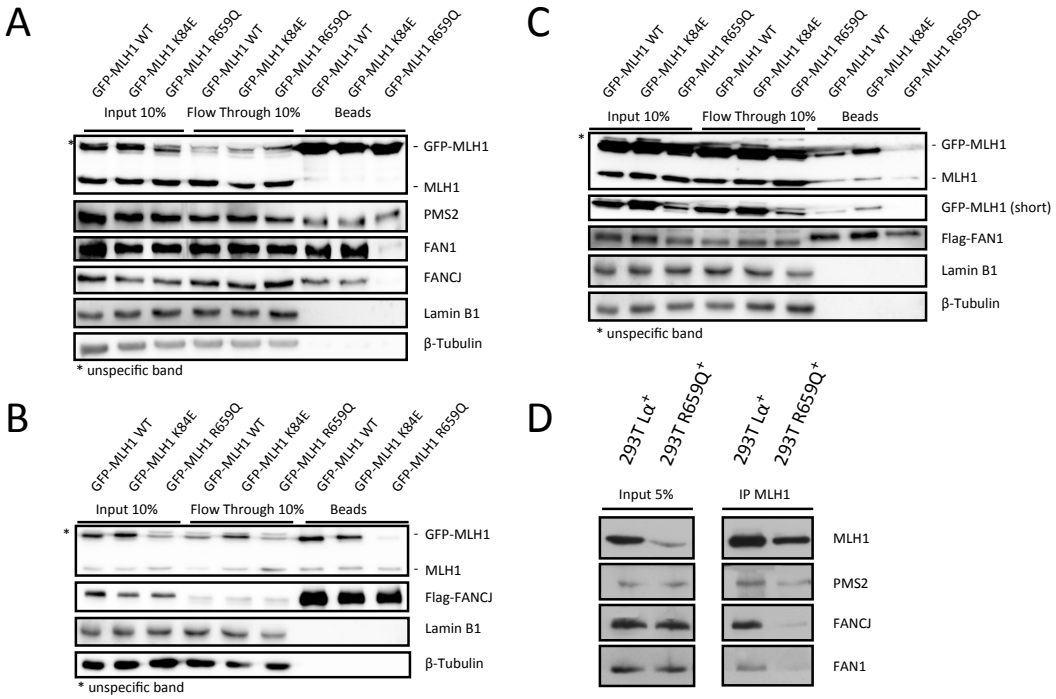


Figure 16: The R659Q mutation impairs binding of MLH1 to FANCJ and FAN1

(**A**) Pulldown assay of GFP-MLH1 variants with GFP-Trap was performed and revealed impaired interaction of the R659Q mutant with FAN1 and FANCJ. (**B**) Flag-FANCJ or (**C**) Flag-FAN1 were transiently expressed in the cell lines and pulldown experiments verified impaired binding of GFP-MLH1 R659Q to FANCJ and FAN1. (**D**) Co-IP of MLH1 in 293T R659Q cell extracts confirmed the reduced binding of MLH1 R659Q to FANCJ and FAN1.

6.5 The R659Q and K84E mutations alter ICL damage response

As the interaction of MLH1 R659Q with FANCJ and FAN1 was found to be impaired, we then wanted to see how the mutant reacts to treatment with ICL-inducing agent mitomycin C (MMC). Cell viability assays were performed after induction of the pAIO system and treatment with MMC. Interestingly, both mutants, the K84E and the R659Q, showed significant resistance against MMC compared to endogenous MLH1 and wt GFP-MLH1 (Figure 17 A). Another hallmark of altered ICL damage response is a G2 arrest after MMC treatment. When the cell lines were treated with MMC either for 24 h with 100 ng or with 25 ng for 96 h, we found a reduced arrest in G2 phase for both conditions in the K84E and the R659Q mutants (Figure 17 B,C). We also performed viability assays with MMC treatment in 293T Lα and L R659Q cells and again found resistance of the R659Q mutant compared to wt MLH1. These observations suggest that the impaired binding of MLH1 R659Q alters the DNA damage response towards ICL lesions. It seems that the lesions are tolerated or repaired differently making the cells more resistant.

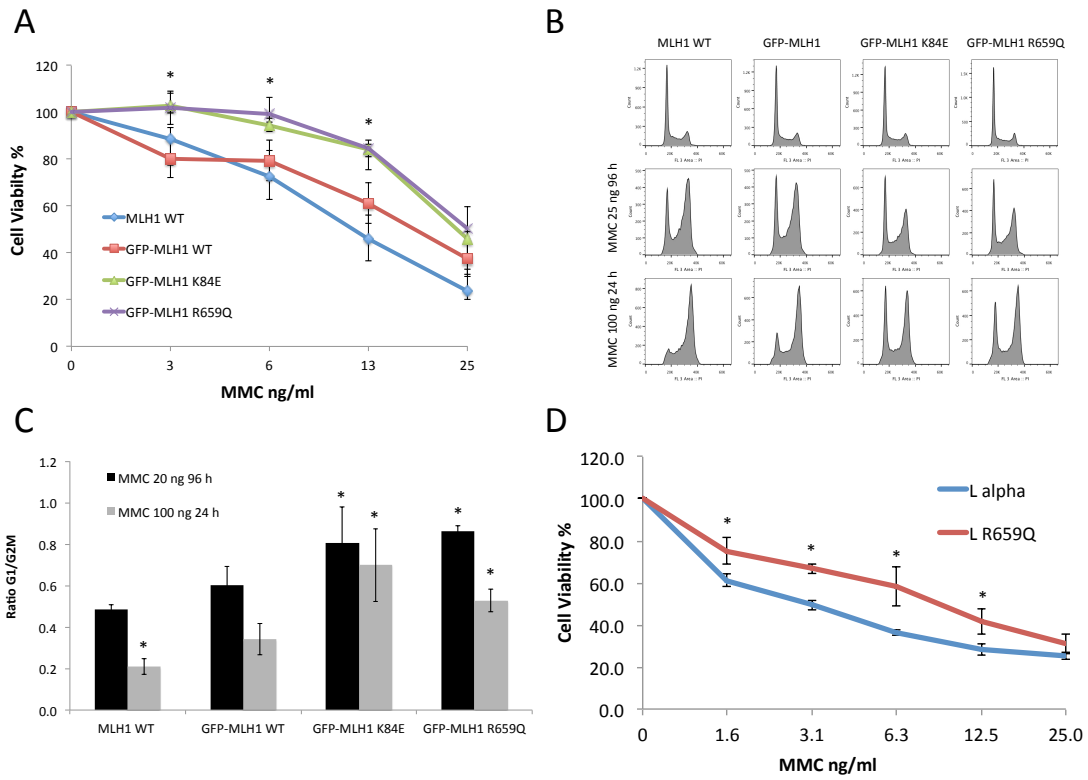


Figure 17: Altered DNA ICL damage response in MLH1 R659Q mutant cells

GFP-MLH1 K84E and R659Q mutants showed resistance against MMC treatment in cell viability assays. Data are represented as mean ± SD from three independent replicates. Asterick indicates significant differences between both mutants K84E and R659Q compared to GFP-MLH1 WT with a p-value below 0.05 according to t-test. (**B**) Cell cycle profile of MLH1 variants after treatment with MMC 25 ng/ml for 96 h or 100 ng/ml for 24 h and release into medium was assessed with PI staining and FACS and the results indicate reduced DNA damage signaling in the K84E and the R659Q mutants as judged by increased portions of cells in the G1 phase and decreased portions in the G2/M phase. (**C**) The ratio of cells in G1 phase to G2/M revealed higher proportions of cells in G1 phase for the K84E and R659Q variants. Data are shown as mean of three replicates ± SD. Asterick indicates p-values below 0.05 compared to GFP-MLH1 WT. (**D**) Cell viability assay with MLH1 variants in the 293T cells confirms the resistance of MLH1 R659Q against MMC treatment. Data are represented from three independent replicates and are displayed as mean ± SD. Astericks indicates significant differences between MLH1 WT and MLH1 R659Q with a p-value below 0.05 according to t-test.

6.6 The pAIO system is reversible

One main advantage of the pAIO system is given by the fact that it is reversible. This was already shown for the 1st generation of pAIO plasmids, when the expression of p125 variants could be stopped simply by washing the cells and release into Dox-free media (130). In this study, we also checked whether the new pAIO system is reversible. When we induced our system by Dox and additionally added siRNA against endogenous MLH1, two washes with PBS and release into Dox-free media was enough to restore endogenous MLH1 and to stop the expression of GFP-MLH1 variants already after 120 h (Figure 18).

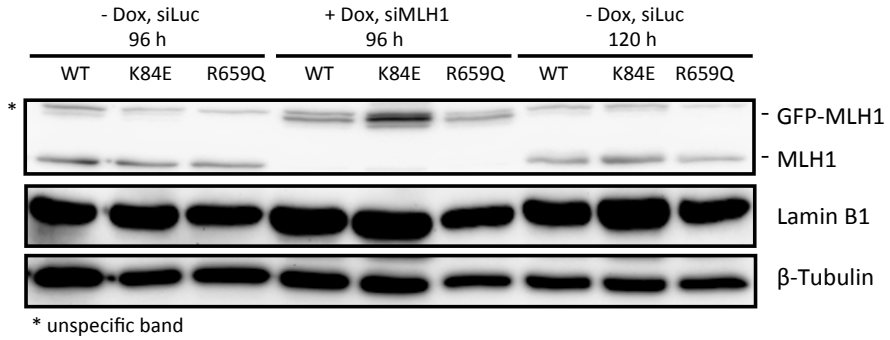


Figure 18: The 2nd generation pAIO system is reversible

Cells with the pAIO system were induced with Dox (+) and treated with siRNAs for 96 h. A fraction was taken for Western Blot analysis and the remaining cells were washed twice with PBS and released into Dox-free medium for 120 h.

6.7 Usage of GFP-tag for interactome studies of MMR factors

In order to study the MMR interactome in different aspects of DNA damage repair, the MMR factors MLH1 and MSH2 were tagged with GFP and eGFP, respectively. For both proteins, the pAIO system was used. For pulldown assays, GFP-Trap was used and the efficiency of the pulldown was assessed by bound interaction partners (Figure 19). For GFP-MLH1, PMS2 and FAN1 could be efficiently enriched, whereas the control PCNA was not found in the eluate. Similarly, MSH2-eGFP could be used to pull down MSH6 and MSH3. As a next step, the bound proteins were sent for LC-MS analysis to identify novel interaction partners of MLH1. In addition, extracts from MNNG-treated cells to induce non-canonical MMR as well as chromatin-enriched extracts were analyzed. Table 3 summarizes the findings of the LC-MS analyses for GFP-MLH1. The main interaction partners of MLH1 including PMS2, PMS1, MLH3 and FANCJ were identified in all three extracts. Except for PRRC2C (proline rich coiled-coil 2C), which was enriched in the non-treated whole cell sample, no strong differences could be observed between the various samples with high values in normalized total spectra (NTS), unique peptide count (UP) and sequence coverage (SC). Two factors could be identified only in the MNNG-treated or chromatin-enriched extracts; KHRSP (KH-type splicing regulatory protein) and NONO (Non-POU domain-containing octamer-binding protein). For MSH2, prominent interaction partners such as MSH6 and MSH3 as well as EXO1 were reliably identified (Table 4). Some interesting factors were found with relatively high values such as WDHD1 (WD Repeat And HMG-Box DNA Binding Protein 1), MCM8 and 9 (Minichromosome Maintenance). However, no obvious changes between MNNG-treated and untreated extracts could be observed.

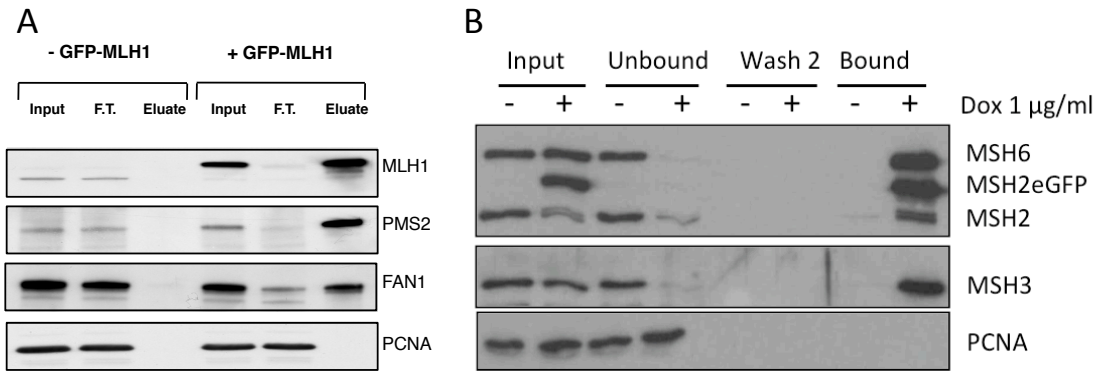


Figure 19: Pulldown of tagged MMR proteins with GFP-Trap

(**A**) GFP-MLH1 or (**B**) MSH2-eGFP was induced in the pAIO system and pulldown assays with GFP-Trap were performed. For MLH1, well-described interaction partners PMS2 and FAN1 were found bound to the beads, while PCNA as a negative control did not bind. For MSH2, the main interactors, namely MSH6 and MSH3, could be found. Again, PCNA did not bind.

	MLH1			MLH1 MNNG			MLH1 Chr		
	NTS	UP	SC	NTS	UP	SC	NTS	UP	SC
MLH1	210	3	66%	104	2	56%	113	1	65%
PMS1	176	31	38%	89	22	26%	97	50	48%
FANCJ	60	3	23%	27	4	15%	50	6	26%
PMS2	54	19	25%	20	17	27%	31	45	36%
PRRC2C	33	16	9.2%	0	0	0%	0	0	0%
MLH3	23	12	11%	4	3	2.3%	16	18	16%
GLI2	21	11	11%	2	2	2.6%	8	9	9%
ARID3B	12	6	17%	7	6	15%	9	10	30%
FMR1	8	4	5.2%	0	0	0%	0	0	0%
SEC22B	8	4	20%	0	0	0%	0	0	0%
BLM	6	3	3.2%	0	0	0.6%	6	7	7.1%
CYLD	6	3	3.5%	1	1	1%	0	0	0%
Exo1	6	3	5.2%	0	0	0%	2	2	4%
FXR2	6	3	5.8%	0	0	0%	0	0	0%
RPS27L	6	1	29%	2	0	15%	3	1	29%
FILA2	4	2	1%	2	2	1%	3	3	2%
NUP153	4	2	1.6%	0	0	0%	2	2	2.2%
ZC3H14	4	2	3.4%	4	3	6.1%	0	0	0%
FAN1	2	1	1%	0	0	0%	4	3	3.8%
MSH3	2	1	1%	0	0	0%	5	5	7.2%
FARSB	0	4	7.3%	5	0	0%	7	8	15%
KHSRP	0	0	0%	4	3	5.1%	8	9	20%
NONO	0	0	0%	0	0	0%	6	7	14%

Table 3: Results of LC-MS of GFP-MLH1 interactome

Interaction partners of GFP-MLH1 identified by GFP-Trap pulldown and LC-MS in untreated cells (MLH1), MNNG-treated (MLH1 MNNG) and chromatin-enriched extracts were sorted according to scores in normalized total spectra (NTS), unique peptides (UP) and sequence coverage (SC).

	MSH2			MSH2 MNNG		
	NTS	UP	SC	NTS	UP	SC
MSH2	139	41	52%	170	43	53%
MSH6	100	38	34%	114	39	34%
MSH3	40	29	36%	62	34	41%
WDHD1	19	15	16%	17	12	12%
EXOI	11	9	16%	20	11	19%
SMRCD	4	3	4.4%	5	4	5.3%
MCM9	4	3	3.5%	7	5	6.3%
CL023	4	3	47%	4	3	40%
1433F	4	3	13%	4	3	13%
MCM8	2	2	3.3%	4	3	5.1%
MLH1	2	2	4.9%	3	2	4%
CCNB1	1	1	3.9%	4	3	9%
COPG2	1	1	1.4%	3	2	2.5%
CD2AP	2	2	3.8%	1	1	2.2%
SET	2	2	8.7%	0	0	0%
CSRP2	2	2	13%	0	0	0%
SPG20	0	0	0%	3	2	4.5%

Table 4: Results of LC-MS of MSH2-eGFP interactome

Interaction partners of MSH2-eGFP identified by GFP-Trap pulldown and LC-MS in untreated (MSH2) and MNNG-treated (MSH2 MNNG) cells were sorted according to scores in normalized total spectra (NTS), unique peptides (UP) and sequence coverage (SC).

6.8 WDHD1 levels decrease upon MSH2 depletion and MNNG treatment

Because WDHD1 was found as a strong binding partner of MSH2 in both approaches (rank 4 directly behind MSH6 and MSH3), we decided to conduct some further experiments with this candidate. The function of WDHD1 [also known as Acidic Nucleoplasmic DNA binding Protein 1 (AND-1) or Chromosome Transmission Fidelity Factor 4 Homolog (CTF4)] still remains enigmatic. Work done so far associates the DNA binding protein with genome maintenance such as signal transduction upon DNA damage, initiation of DNA replication and a role in homologous DNA double strand break repair (144-147). Depletion of WDHD1 was shown to increase DNA damage, prolong S phase, induce cell death in cancer cells and decrease repair of I-SceI-driven double-strand breaks repair. Moreover, knock down of WDHD1 impaired the formation of the CMG complex (Cdc45,

MCM2-7, GINS), which is believed to play an important role for DNA replication fork formation and progression.

As a first experiment, HEK 293 cells were depleted of MSH2 and protein levels of WDHD1 were addressed. Interestingly, the protein levels of WDHD1 decreased after 72 h and 96 h (Figure 20 A). As WDHD1 has been associated with maintenance of genome stability and non-canonical MMR can be induced by MNNG, we checked for recruitment of WDHD1. Surprisingly, unlike MMR factors, WDHD1 levels decreased upon MNNG treatment in the nucleus in HEK 293 as well as in U2OS cells (Figure 20 B). For U2OS cells, a decrease in cytoplasmic fraction was also observed. As illustrated in Figure 20C, depletion of MSH2 in MNNG cells did not rescue or further decrease WDHD1 levels, which implies that the decrease upon MNNG treatment is independent of MSH2.

Recently, Chen and colleagues also identified WDHD1 as a strong interaction partner of MSH2 in an AP-MS approach and could verify the interaction with Co-IP. Moreover, they found that WDHD1 depletion makes cells resistant towards 6-thioguanine (6-TG) treatment, as does MSH2 depletion (148).

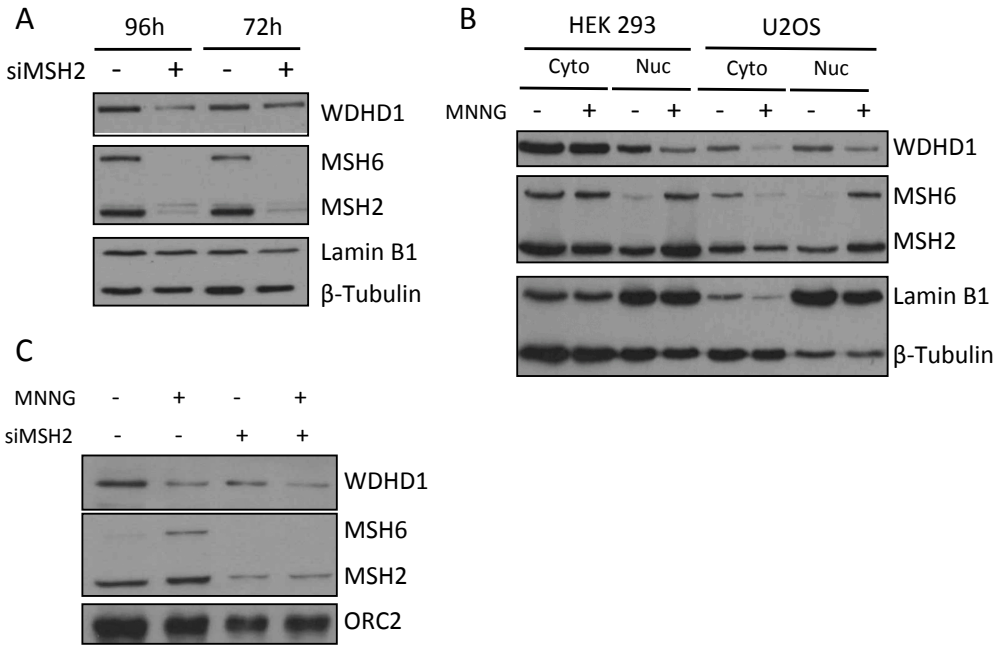


Figure 20: Decrease of WDHD1 protein levels upon MSH2 depletion and MNNG treatment

(**A**) HEK 293 cells were treated with siRNA MSH2 (+) or siLuc (-) for indicated time periods. Western Blot analyses revealed decreased levels upon MSH2 knockdown. (**B**) HEK 293 and U2OS cells were treated with 10 mM O6-BG for 1 h and 5 µM MNNG for 3 h and cytoplasmic (Cyto) and nuclear (Nuc) proteins were isolated. For both cell lines, WDHD1 levels decreased in the nucleus upon MNNG treatment. (**C**) Nuclear proteins of MNNG-treated and siRNA-treated cells were analyzed and revealed no influence of MSH2 on the reduction of WDHD1 levels upon MNNG treatment.

6.9 Proximity-dependent biotin labeling using BioID

The applied AP-MS approach using GFP or eGFP-tagged MMR proteins and GFP-Trap for pulldowns revealed WDHD1 as an interesting new interaction partner of MutSα and MutSβ, although our initial aim to develop a technique that could be used to analyze the MMR interactome before and after DNA damage could not be addressed. Therefore, we decided to explore other methods. The first one we tested is a technique called BioID. A mutated biotin ligase (BirA* R118G) fused to a protein of interest produces reactive biotinyl-5'-AMP from biotin, which can bind to lysine residues of any protein in close proximity (128).

6.9.1 BirA*-MLH1 is not active, but PMS2-BirA* is

BirA* was introduced into the pAIO GFP-MLH1 plasmid by replacing the GFP-tag and stable cell lines were generated. When induced by Dox, only low levels of BirA*-MLH1 could be observed in total cell extracts and essentially no signal was detected in the nucleus (Figure 21 A). Moreover, treatment with MNNG did not increase nuclear BirA*-MLH1, while endogenous MLH1 was recruited into the nucleus. When we tested the enzymatic activity of BirA*-MLH1 by checking for biotinylated proteins using a Strep-HRP, we did not observe any biotinylation (Figure 21 B). The positive control, transiently transfected BirA*-Myc, however, showed very strong biotinylation of endogenous proteins. We also tested a PMS2-BirA* construct and could detect biotinylation of endogenous proteins.

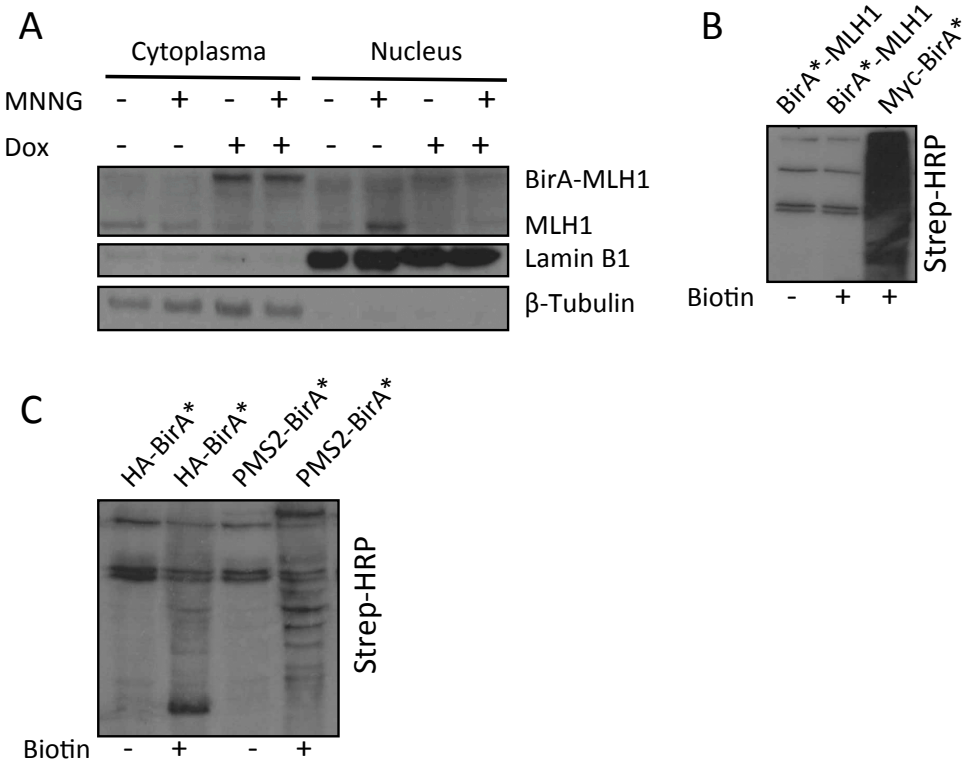


Figure 21: PMS2-BirA*, but not BirA*-MLH1 is active

(**A**) Cytoplasmic and nuclear extracts of cells stably expressing BirA*-MLH1 with (+) or without (-) MNNG treatment revealed erroneous localization of BirA*-MLH1 and no recruitment to the nucleus after MNNG treatment. (**B**) Western Blot with Strep-HRP showed no biotinylation of endogenous proteins in BirA*-MLH1 cells. (**C**) Transiently expressed PMS2-BirA* can biotinylate endogenous proteins.

6.10 Proximity-dependent biotin labeling using Apex

As the Bio-ID method did not work for MLH1, we decided to try an approach called APEX. APEX is an engineered ascorbate peroxidase from soybean that generates highly-reactive phenol radicals that can bind to electron-rich amino acids such as tyrosine, tryptophan, cysteine or histidine. The reaction is catalyzed by H_2O_2. When biotin-phenol is used, endogenous proteins in close proximity are labeled with biotin-phenol and can later be isolated on streptavidin-coated resins (129).

6.10.1 APEX2-MLH1 is active and behaves as endogenous MLH1

We generated an APEX2-MLH1 cell line using the pAIO system. Instead of APEX, we used APEX2, which has increased activity due to an additional A134P mutation (132). As shown in Figure 22 A, APEX2-MLH1 was highly expressed in the cytoplasm and the protein levels in the nucleus were also stronger than endogenous MLH1. Upon MNNG treatment, APEX2-MLH1 was recruited to the nucleus. When we treated the cells expressing APEX2-MLH1 with biotin-phenol followed by H_2O_2, we could detect strong labeling of endogenous proteins in the cytoplasm as well as in the nucleus (Figure 22 B). The labeling in the nucleus was stronger than in the cytoplasm.

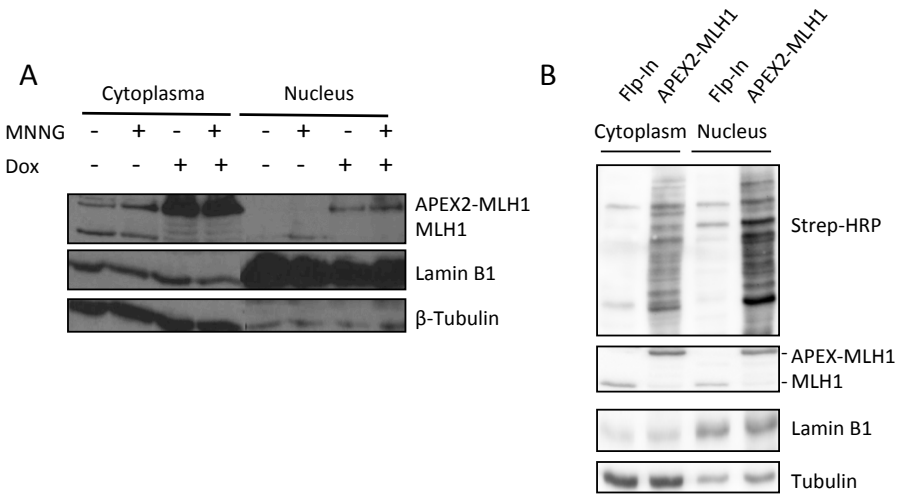


Figure 22: APEX2-MLH1 is recruited to the nucleus and is enzymatically active

(**A**) Western Blot of MNNG (+) and untreated cells (-) showed enhanced recruitment of APEX2-MLH1 to the nucleus after DNA damage. (**B**) Cytoplasmic and nuclear proteins of cells treated with biotin-phenol and H_2O_2 reveal labeling with biotin as judged by staining with streptavidin-HRP.

As a next step, we used streptavidin agarose beads to pull down biotinylated proteins, which are supposed to be interactors of MLH1. As shown in Figure 23 A, the pulldown was clear as we only found endogenously biotinylated proteins in the control. In the APEX-MLH1 cells, many bands appeared, which represent proteins that were biotinylated through the APEX reaction and are potential interactors of MLH1. To confirm that interactors of MLH1 could be enriched, a Western Blot was performed and probed with antibodies for well-known MLH1 interaction partners (Figure 23 B). In addition to the main binding partner PMS2, we could detect FAN1 and MSH2. Together, these experiments demonstrated that APEX-MLH1 is enzymatically active and interaction partners of MLH1 can be labeled with biotin-phenol and subsequently pulled down with streptavidin beads.

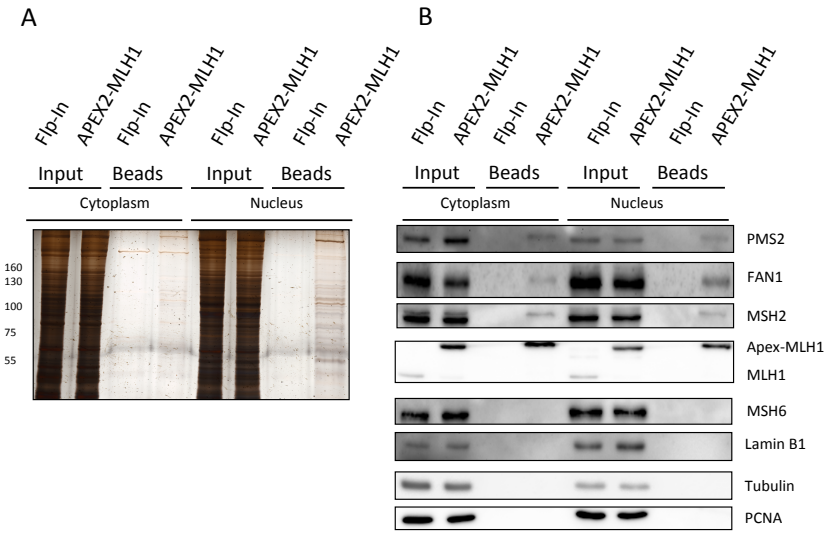


Figure 23: Enrichment of MLH1 interactors after APEX2-MLH1 activation

(**A**) Proteins purified with streptavidin agarose were made visible by silver staining. In the negative control (Flp-In), only endogenous biotinylated proteins are visible, while in the extracts of APEX2-MLH1 cells, several protein bands appeared. (**B**) Western Blot of pulldowns from A revealed biotinylation of known MLH1 interaction partners PMS2, FAN1 and MSH2.

6.10.2 MLH1 interactome using APEX approach

Having shown that APEX2-MLH1 is active, we went on to study the MLH1 interactome with this system. We performed LC-MS analyses of two replicates of each condition and fraction. Flp-In cells were used as control. As shown in Figure 24, over 1,300 proteins were detected in each APEX2-MLH1 sample and no more than 900 in all controls. Moreover, 1,381 proteins were identified in APEX-MLH1 samples alone. We also observed a difference between cytoplasmic and nuclear extracts. 525 factors were detected in both fractions, but 337 proteins were exclusively found in the cytoplasm and 519 in the nucleus, respectively. The sequence coverage of MLH1 was rather high with about 48 - 68%. As a quality control and in order to validate our results, the data were aligned to protein interaction databases such as IntAct (149). Many known and well-established interactors could be seen in our approach including PMS1, PMS2 and Exo1. However, the data on some factors were quite unexpected. For instance, PMS2 was identified with high spectra count in the cytoplasm but not in the nucleus. MSH2 and FAN1, which were clearly identified by Western Blot did not match the threshold criteria and are absent in the list. For MSH2, 20 and 13 peptides could be identified in the cytoplasm and nucleus in the APEX2-MLH1 cells, respectively, but also 7 and 3 peptides in the control. FAN1 was not assigned at all.

A

Samples	proteins id	MLH1 coverage (%)	streptavidin
APEX_MLH1_nuclear_2	1539	68.4	58.1
APEX_MLH1_nuclear_1	1365	56.7	65
APEX_Flp_nuclear_2	848	11.8	58.1
APEX_Flp_nuclear_1	841	9.3	55.6
APEX_MLH1_cytoplasmic_2	1701	59.1	58.1
APEX_MLH1_cytoplasmic_1	1687	48.1	62.5
APEX_Flp_cytoplasmic_2	630	ND	55.6
APEX_Flp_cytoplasmic_1	793	ND	48.8

B

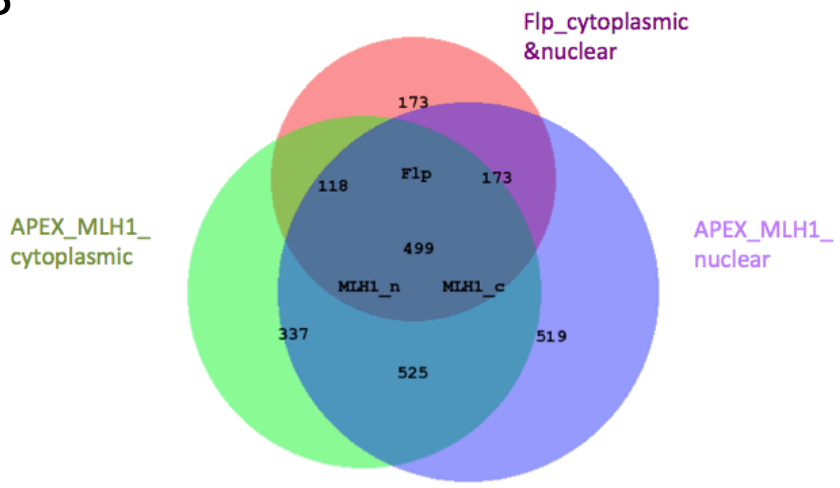


Figure 24: LC-MS APEX2-MLH1 nuclear and cytoplasmic extracts

(**A**) Cells expressing APEX2-MLH1 and 293 Flp-In cells were treated with biotinphenol and H_2O_2 followed by the extraction of cytoplasmic and nuclear proteins, streptavidin pulldown and LC-MS. The total numbers of identified proteins and sequence coverage of MLH1 and streptavidin in each fraction and replicate are shown. (**B**) Venn diagram of identified proteins reveals that 1,381 proteins were identified exclusively in APEX2-MLH1 samples, 525 of which were detected in the cytoplasmic and nuclear fraction. 337 proteins were found in the cytoplasmic fraction and 519 in the nuclear fraction, respectively.

Gene	Spectra APEX2-MLH1 Cytoplasm	Spectra APEX2-MLH1 Nucleus	Flp-In Cytoplasm/ Nucleus
MSH6	41|52	47|50	4|5|13|10
CKB	41|49	0|0	0|0|5|2
FANCD2	32|35	9|11	0|0|0|0
ALDOA	25|28	0|0	5|0|0|0
PMS1	21|32	7|7	0|0|0|0
ATM	19|29	8|7	0|0|0|0
PMS2	19|22	0|0	0|0|0|0
RUVBL2	18|17	22|22	3|4|2|3
RUVBL1	19|14	21|21	2|2|2|2
FANCJ	12|14	0|0	0|0|0|0
MSH3	12|11	19|22	0|0|2|3
NPP2R1A	10|12	0|0	3|2|0|0
PSMA1	11|9	0|0	0|0|0|0
MMS19	10|9	0|0	2|0|0|0
SNW1	7|10	8|8	0|0|0|0
NBN	8|7	7|6	0|0|0|0
ANXA6	3|7	0|0	0|0|0|0
BRCA1	2|6	6|6	0|0|0|0
RAD23B	3|4	0|0	0|0|0|0
MLH3	5|2	0|0	0|0|0|0
PYGB	2|4	0|0	0|0|0|0
CUTA	3|3	0|0	0|0|0|0
SQSTM1	3|2	0|0	0|0|0|0
PPP2R2A	2|3	0|0	0|0|0|0
EXO1	2|3	6|6	0|0|0|0
NDGR1	2|2	0|0	0|0|0|0
NONO	0|0	164|180	7|9|40|39
ZC3H11A	0|0	25|27	0|0|4|3
PRKDC	0|0	608|637	64|64|208|189
YLPM1	0|0	67|89	0|0|13|7
BARD1	0|0	4|4	0|0|0|0
BLM	0|0	16|18	0|0|0|2
RFC1	0|0	29|33	0|0|8|10

Table 5: Known MLH1-interacting proteins identified by APEX2-MLH1 approach

Proteins identified in LC-MS (Figure 24) were scored by total spectra count and submitted to the IntAct database to align the hits against known MLH1 interacting proteins (149). The results are listed according to their score in the cytoplasmic fraction.

6.10.3 Differential MLH1 interactome after MNNG treatment

Despite the contradicting results from the Western Blot analyses and the LC-MS regarding FAN1 in particular, we went on to study the MLH1 interactome after DNA damage. First, we decided to treat the cells with MNNG to study the MMR interactome after alkylation damage and induction of non-canonical MMR. Cells expressing APEX-MLH1 were treated with O6-BG for 1 h, followed by treatment with 10 µM MNNG or DMSO for 3 h. Biotin-phenol was added 2 h after addition of MNNG. When we analyzed biotinylated proteins by LC-MS, we found several factors enriched in the MNNG samples compared to DMSO treated (Table 6). We found several factors involved in translation such as RPN1, TUFM and RPS18 as well as proteins associated with protein folding and membrane trafficking including CANX and SSR4. We also found proteins associated with DNA repair such as WDR8, USP22 and BCCIP. Interestingly, WDR18 and BCCIP appear as interactors in a string network created with the identified proteins (Figure 25). In addition to factors that appeared only after MNNG treatment as potential interactors of APEX2-MLH1, we could identify some proteins that disappeared after MNNG treatment (Table 7). Because BCCIP (BRCA2 and CDKN1-A interacting protein) has previously been shown to interact with RAD51 as well as BRCA2 and has been suggested to play a role in HR, we decided to do some follow-up experiments (150). Sensitivity against alkylation damage is a very good readout for proteins involved in MMR and because BCCIP was found as a potential co-factor of MLH1 after MNNG treatment, we decided to check for cell survival after BCCIP single and BCCIP/MLH1 double knockout. As expected, MLH1 knockdown made the cells resistant against MNNG (Figure 26). The phenotype of BCCIP depleted cells resembled the one of wt cells. Moreover, cells lacking both, MLH1 and BCCIP, behaved as MLH1-deficient cells. We could neither observe any effect after MNNG treatment of BCCIP knockdown alone nor additional or rescue effects of BCCIP on MLH1 depleted cells.

Gene	Spectra MNNG	Spectra DMSO	AvgP	BFDR
RPN1	8|8|4	0|0|0	1	0
PRDX4	4|5|6	0|0|0	1	0
CANX	7|8|3	0|0|0	0.99	0
CPNE3	4|3|3	0|0|0	0.98	0
PTBP2	4|4|2	0|0|0	0.95	0.01
TUFM	2|4|4	0|0|0	0.95	0.01
USP22	2|4|4	0|0|0	0.95	0.01
INTS5	4|2|3	0|0|0	0.94	0.03
WDR18	3|3|2	0|0|0	0.94	0.03
RPS18	2|4|2	0|0|0	0.91	0.03
HSD17B12	2|4|2	0|0|0	0.91	0.03
SPDL1	2|3|2	0|0|0	0.9	0.05
ZFP91	2|2|3	0|0|0	0.9	0.05
BCCIP	3|2|2	0|0|0	0.9	0.05
EIF2B3	3|2|2	0|0|0	0.9	0.05
SSR4	2|2|3	0|0|0	0.9	0.05
LBR	6|6|7	0|2|2	0.9	0.04
CAPZA2	3|2|2	0|0|0	0.9	0.05
TRMT112	3|2|2	0|0|0	0.9	0.05
KRT77	2|3|2	0|0|0	0.9	0.05
ZNF11	2|3|2	0|0|0	0.9	0.05
RPLP1	3|2|2	0|0|0	0.9	0.05

Table 6: Factors with increased APEX2-MLH1 biotinylation after MNNG treatment

Cells expressing APEX2-MLH1 were treated with 10 µM O6-BG for 1 h followed by 10 µM MNNG or DMSO for 3 h before the activation of APEX2. The experiment was done in triplicates and the factors were scored according to the average probability using total spectra count. In addition, the Bayesian false discovery rate (BFDR) is shown.

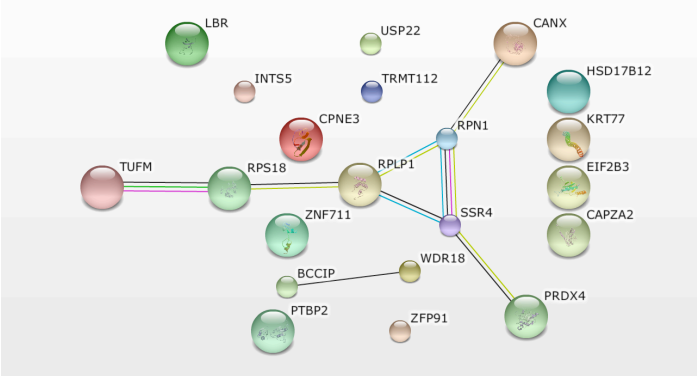


Figure 25: String network of MLH1-interacting proteins after MNNG treatment

Data from Table 6 were used to create a string network for functional analysis (151).

Gene	Spectra DSMO	Spectra MNNG	AvgP	BFDR
HOXB6	2|3|3	0|0|0	0.87	0
PPP4R2	2|3|2	0|0|0	0.8	0.13
ARID4A	2|2|2	0|0|0	0.73	0.16
TFG	2|2|2	0|0|0	0.73	0.16
SUDS3	2|2|2	0|0|0	0.73	0.16
MLLT3	2|2|2	0|0|0	0.73	0.16
MEIS2	2|2|2	0|0|0	0.73	0.16

Table 7: Factors with decreased APEX2-MLH1 biotinylation after MNNG treatment
Cells expressing APEX2-MLH1 were treated with O6-BG (10 µM) for 1 h followed by 10 µM MNNG or DMSO for 3 h before activation of APEX2. The experiment was done in triplicates and the factors were scored according to the average probability using total spectra count. In addition, the Bayesian false discovery rate (BFDR) is shown.

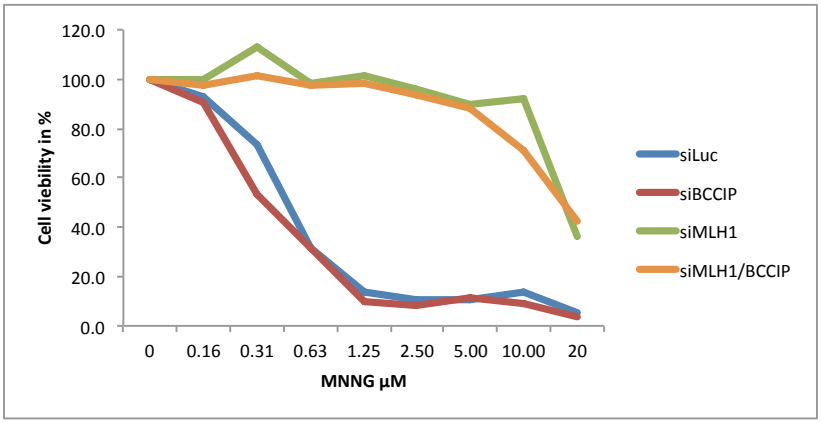


Figure 26: BCCIP knockdown does not alter cell sensitivity towards MNNG
293 cells were treated with siRNAs against MLH1 and BCCIP for 24 h and seeded in 96-well plates. After treatment with O6-BG and MNNG, the cells were incubated for 96 h and the viability was measured. Data points are the average of 3 replicates.

6.10.4 MLH1 interactome after MMC treatment

In addition to MNNG treatment, we wanted to know how the interactome of MLH1 might change after treatment with ICL inducing MMC. APEX2-MLH1 expressing cells were treated with MMC or DMSO and biotinylated proteins from nuclear abstracts were analyzed using LC-MS. As shown in Table 8, several factors reacted on the MMC treatment and can be considered as potential MLH1 interactors after MMC damage. In addition to numerous proteins that have been

associated with functions in development, pre-mRNA splicing, transcription or protein trafficking, factors with functions in DNA repair could also be identified. RPA2 is a well-known factor of MMR and protects single stranded DNA (13). Moreover, Excision Repair Cross-Complementation Group 6 (ERCC6) was found, which has been associated with transcription-coupled DNA repair (152). In contrast to several new potential interactions of MLH1 in response to MMC treatment, only a few factors were lost. The proteins that we could only detect in DMSO-treated cells are summarized in Table 9. Interestingly, two of the three proteins are associated with epigenetic gene regulation (153,154).

Gene	Spectra MMC	Spectra DMSO	AvgP	BFDR
CTNNBL1	15|12|16	3|2|4	0.99	0.01
TEX10	4|3|5	0|0|0	0.99	0
CSNK1E	4|3|2	0|0|0	0.96	0.01
RPA2	5|6|5	0|1|1	0.95	0.02
PHC2	2|2|4	0|0|0	0.94	0.03
SSB	2|2|2	0|0|0	0.91	0.04
DGKH	2|2|2	0|0|0	0.91	0.04
SLC25A6	1|4|5	0|0|0	0.86	0.05
KPNB1	3|3|1	0|0|0	0.85	0.06
TFCP2	3|1|3	0|0|0	0.85	0.06
COASY	2|4|4	1|0|0	0.84	0.08
NCAPG	4|2|1	0|0|0	0.83	0.08
NUP155	4|2|1	0|0|0	0.83	0.08
RBM42	1|4|2	0|0|0	0.83	0.08
PHF6	3|1|2	0|0|0	0.82	0.1
WBSCR22	2|1|3	0|0|0	0.82	0.1
BOP1	2|2|1	0|0|0	0.8	0.11
ERCC6	2|1|2	0|0|0	0.8	0.11
EYA3	2|2|1	0|0|0	0.8	0.11
MAGOH	1|2|2	0|0|0	0.8	0.11

Table 8: Factors with increased APEX2-MLH1 biotinylation after MMC treatment
Cells expressing APEX2-MLH1 were treated with MMC (25 µM) for 1 h before activation of APEX2. The experiment was done in triplicates and the factors were scored according to the average probability using total spectra count. In addition, the Bayesian false discovery rate (BFDR) is shown.

Gene	Spectra DMSO	Spectra MMC	AvgP	BFDR
KMT2B	4|4|3	0|0|0	0.95	0
KPNA1	3|5|3	0|0|0	0.93	0.05
EZH2	3|2|2	0|0|0	0.69	0.06

Table 9: Factors with decreased APEX2-MLH1 biotinylation after MMC treatment
Cells expressing APEX2-MLH1 were treated with MMC (25 μM) for 1 h before activation of APEX2. The experiment was done in triplicates and the factors were scored according to the average probability using total spectra count. In addition, the Bayesian false discovery rate (BFDR) is shown.

6.10.5 APEX2-GFP labels proteins unspecifically

When APEX2-MLH1 was used to find new co-factors of MLH1, many known interactors could be identified. However, more than 1,300 proteins were found to be enriched in the APEX2-MLH1 samples compared to the control. Considering that more than 1,300 proteins account for about 5% of the total human proteome and are too many to be followed up step-by-step, we decided to introduce another control to reduce the number of new potential MLH1-interacting proteins. Hence, we generated an APEX2-GFP fusion protein and made stable cell lines again using the pAIO system. After induction of APEX2-GFP, we treated the cells with biotin-phenol and H_2O_2 and found that APEX-GFP is active in labeling endogenous proteins with biotin-phenol in the cytoplasm as well as in the nucleus (Figure 27 A). When we compared the activity of APEX2-GFP with APEX2-MLH1, we found that APEX2-GFP is slightly more active in the cytoplasm than APEX2-MLH1 but has a similar activity in the nucleus (Figure 27 B). The activity of APEX2-GFP was stronger in the nucleus than in the cytoplasm. The same was observed for APEX2-MLH1. Fluorescence microscopy was used to check for the distribution of APEX2-GFP within the cell. According to the GFP-signal, APEX2-GFP localized equally to the cytoplasm and to the nucleus (Figure 27 C).

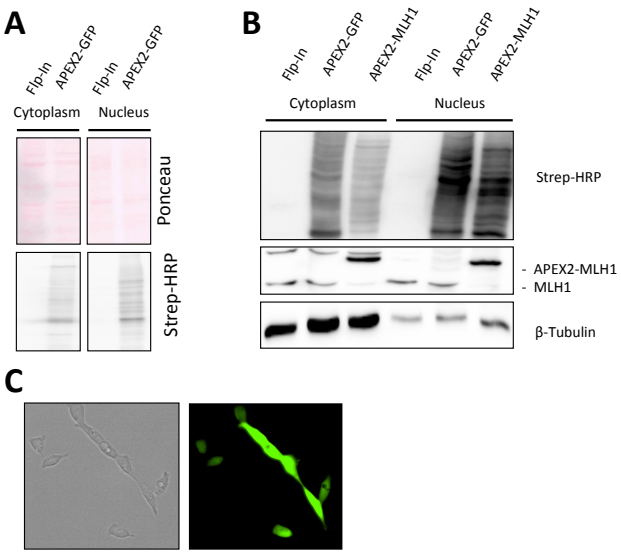


Figure 27: APEX2-GFP is active in the cytoplasm and the nucleus

(**A**) Parental 293 Flp-In cells and APEX2-GFP expressing cells were treated with biotin-phenol and H_2O_2 which led to biotinylation of endogenous proteins, as judged by incubation with streptavidin-HRP. (**B**) Flp-In APEX2-GFP and APEX2-MLH1 cells were treated with biotin-phenol and H_2O_2 to check for APEX2 activity. (**C**) White light (left) and UV light (right) show the distribution of APEX2-MLH1 all over the cell.

As a next step, a pulldown of proteins biotinylated by APEX2-GFP and APEX-MLH1 was performed to check for unspecific biotinylation of MLH1-binding proteins by APEX2-GFP. Unfortunately, the Western Blot revealed that APEX2-GFP seems to label all proteins. Every protein we looked at, even including the loading control β-Tubulin, was found to be biotinylated (Figure 28). Moreover, some interactors of MLH1 were labeled even stronger by APEX2-GFP than by APEX-MLH1, such as FANCJ, FAN1 and MSH2 in the cytoplasmic fraction. Also, MLH1 was found to be biotinylated by APEX2-GFP. Only PMS2 was significantly more enriched in the APEX2-MLH1 samples. As a next step, we had planned LC-MS analyses to characterize the MLH1 interactome using the APEX approach. For obvious reasons, we did not continue and started to develop other controls such as an APEX2-TDG fusion protein.

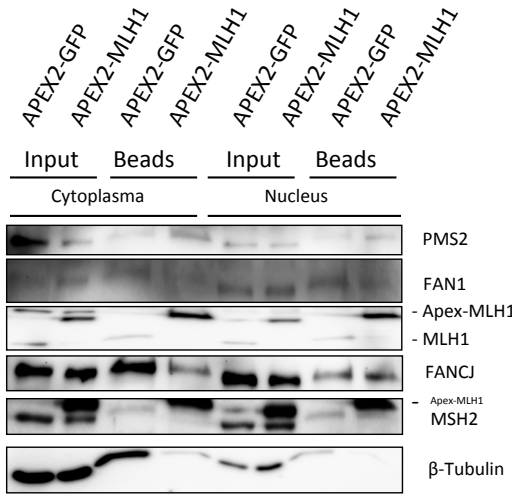


Figure 28: APEX2-GFP labels MLH1 interactors and β-Tubulin

Cells expressing either APEX2-GFP or APEX2-MLH1 were treated with biotin-phenol and H_2O_2 and cytoplasmic and nuclear proteins were extracted. After pulldown with streptavidin resins, the bound proteins were analyzed by Western Blot. All tested proteins were biotinylated by APEX2-GFP.

7 Discussion

7.1 The pAIO system – a powerful method to study gene variants

The pAIO protein replacement system has been developed to study genotype-phenotype correlations and in particular missense mutations. The original pAIO system consisted of two main features: an shRNA expression cassette to knock down the endogenous protein and a cDNA expression cassette for transcription of a protein variant. Importantly, both cassettes are inducible by Dox. The system has been used successfully to replace the catalytic subunit of DNA polymerase pol-δ p125 (*POLD1*) with a flag-tagged variant of p125 with either wt, error-prone or proofreading deficient variants (130). Many different experiments and assays could be performed by the simple addition of Dox to induce the system instead of repeated transfection of cDNA variants and siRNAs. However, the generation of the final cell lines was work-intensive and required many individual cloning steps as well as picking and testing of several individual single cell clones.

To improve our system, we aimed to combine it with well-established fast cloning methods and a system to overcome the tedious work of single cell clone analysis. Hence, we decided to introduce the Gateway system into our plasmid to speed up the integration of cDNAs into the pAIO plasmid. Using this approach, the integration of cDNA variants from Entry clones became more rapid and reliable. Another advantage of the Gateway system is clearly given by the easy generation of gene variants such as missense mutations. Instead of mutagenesis of the whole final plasmid, only the entry clone needs to be mutated and the variant can then be cloned into the final pAIO destination vectors. This saves a lot of time and resources regarding sequencing, because only the cDNA sequence in the Entry vector needs to be checked instead of the final plasmid. In order to advance the generation of stable cell lines, we made use of the Flp Recombination Target (FRT) system. Using this system, picking of single cell clones in Flp-In 293 T-REx cells became dispensable and it was possible to generate homogenous bulk cell lines in less than three weeks. Although the CMV

promoter used in our plasmid led to a p125 expression very similar to endogenous levels, it was too strong for GFP-MLH1. In order to reduce the expression of the CMV promoter, we developed strategies based on PCR to shorten the CMV promoter and thereby the expression levels of the cDNA. An overview of the use and improvements of the 2nd generation pAIO system is shown in Figure 29.

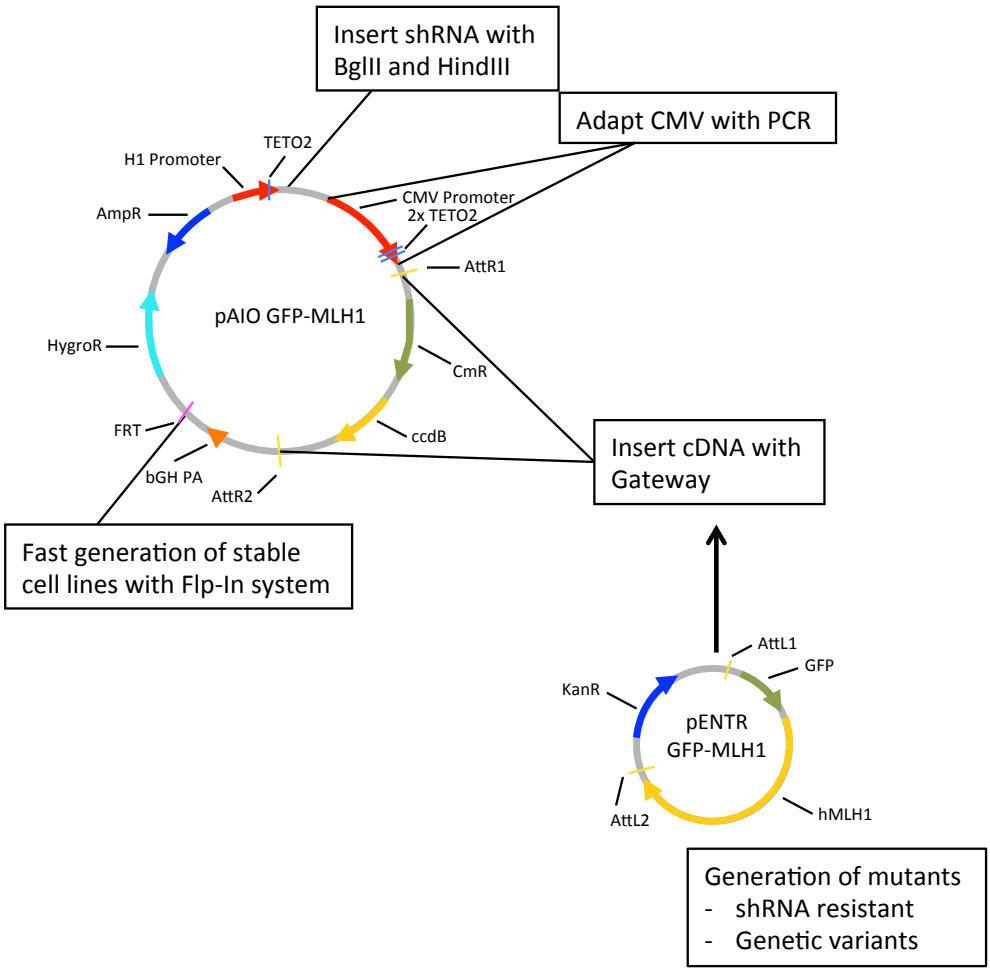


Figure 29: Scheme of the 2nd pAIO protein replacement system

The pAIO vector contains an H1 promoter for shRNA expression and a CMV promoter for cDNA expression, both inducible due to tet binding sites. The shRNA oligonucleotide is easily integrated into the plasmid with BglII and HindIII digestion. The cDNA of interest is recombined into the vector via AttR sites for Gateway cloning. With PCR techniques, the CMV promoter can be adapted to the desired expression. For the generation of stable cell lines, the FRT site is used in Flp-In cells and stable cells are selected with hygromycin. The usage of entry plasmids before Gateway recombination provides a reliable method to generate shRNA-resistant cDNAs and mutant variants of the gene of interest.

7.2 The pAIO system as an alternative to the CRISPR/Cas9 system

The CRISPR/Cas9 (Clustered regularly-interspaced short palindromic repeats/CRISPR associated protein 9) system provides a very powerful tool to specifically knock out genes in any cell. In principle, a single guide RNA (sgRNA) binds homology-dependently to a specific sequence in the genomic DNA utilized by its crRNA (CRISPR RNA) domain and recruits Cas9 with its tracrRNA (trans-activating crRNA) domain. The nuclease activity of Cas9 creates a DNA DSB, which is subsequently repaired either by NHEJ or by homology-directed repair (HDR) (155,156). In both cases, the aim is to disrupt the gene sequence leading to the loss of the functional gene product. Once a gene has been knocked out, it can be replaced by insertion of a cDNA copy similar to our pAIO plasmid. In the same manner, different cDNA variants such as missense mutations can be integrated into the knockout cell. Another way of creating a cell line with a specific mutation is provided by the HDR approach. In this case, a DNA DSB is induced and a plasmid containing the mutated sequence is used as a template for homologues repair resulting in a cell line with a mutated gene.

The CRISPR/Cas9 system has one big advantage over the pAIO system. While our pAIO system is based on RNAi, which leads only to a knockdown of the target gene, the CRISPR/Cas9 system results in a complete gene knockout. For some target genes, an shRNA-mediated knockdown is very efficient and basically no protein can be detected. For instance, no visible p125 after knockdown with the 1st generation pAIO plasmid was left (130). In contrast, the knockdown of MLH1 with shRNA or siRNA alone was strong, but longer exposure times revealed more than 10% of residual endogenous MLH1, which can many MLH1 functions (72). Only the combination of both shRNA from the pAIO plasmid and transient transfection of siRNA reduced MLH1 way below 10%. Moreover, it is believed that RNAi-mediated gene knockdown results in more off-target effects than the use of CRISPR/Cas9. As we replace the endogenous protein with a tagged cDNA variant in our system, we expected to see no significant differences between the non-induced situation and the replacement with the wt cDNA. This is exactly what we observed in most experiments meaning that the chance of artifacts due to off-target effects is rather low. Only in the cell viability assay after MNNG

treatment did we observe that wt GFP-MLH1 cells were slightly more resistant than non-induced cells with endogenous MLH1. This could hint towards an off-target effect, but it is more likely that GFP-MLH1 is slightly less efficient than MLH1 in addressing MNNG-induced DNA lesions. Another big advantage of the CRIPR/Cas9 system is given by the HDR approach. If the desired mutation is induced within the genome by HDR, the gene is still expressed from its own promoter resulting in normal expression levels and still underlies endogenous gene regulation. Although this has not been a problem for the replacement of p125, the expression of MLH1 from the wt CMV promoter was too strong, which was, however, overcome by adapting the promoter.

Clearly, there are some strong advantages of the CRISPR/Cas9 system over the pAIO system but it also has some disadvantages. First, a knockout is not possible for all genes. Some genes, mainly structural genes, are essential and a complete knockout ultimately leads to cell death, a fact that makes it impossible to create cell lines for later replacement with cDNA variants. The pAIO system, however, is inducible, which allows for simultaneous replacement of the endogenous gene with a cDNA variant. Thus, even essential genes could be studied with the pAIO system, because the knockdown is only induced after Dox treatment which allows for the selection of stable cell lines. Even the HDR approach might not help here, because missense mutations that abolish the function of an essential gene will also result in cell death. Second, depending on the efficiency of the gene knockout with CRISPR/Cas9, it can be much work to screen many single cell clones for every cDNA variant. Although this amount of work can be reduced by integrating a resistance gene using HDR, it is still much more work than using the pAIO system. Third, we have shown that very reliable bulk cell lines can be generated with Flp-In 293 T-REx cell lines. In the case of CRISPR/Cas9, several single cell clones ought to be used for every experiment to exclude clonal effects.

Another limitation of the CRISPR/Cas9 system concerns gene copy numbers. For most genes, cells carry two copies of a single gene and both need to be knocked out or replaced. However, cancer cell lines with altered karyotypes such as U2OS cells may not only have two but several copies of the target gene and all of them

have to be hit. Besides changes of copy numbers due to cancerous transformation, there are genes that have naturally more than two copies. Histone genes have 10 - 20 gene repeats in higher eukaryotes and up to 100 in *drosophila melanogaster* (157,158). These numbers make it nearly impossible to knock out all of them step by step. An RNAi based approach, however, might provide a solution as it does not target the DNA but the transcripts.

7.3 Combination of pAIO and CRISPR/Cas9 to overcome drawbacks

The pAIO and the CRIPR/Cas9 system have their advantages and disadvantages. While the biggest benefit of the CRISPR/Cas9 system is clearly the complete knockout of a gene, the strongest argument for the pAIO system is provided by the fact that it is inducible and reversible. The Cas9 protein has been engineered to be used for many other applications besides the induction of DNA DSBs. One interesting development of the Cas9 toolbox is the fusion of a nuclease dead Cas9 protein, which is now converted into a sequence-specific DNA-binding protein, with transcriptional repressors such as KRAB (Krüppel associated box) (159). The chromatin modifier KRAB recruits histone deacetylases and methylases such as Mi-2α and SETDB1, respectively, which bring about an epigenetic change and thus can shut down the transcription of the target gene (160,161). This approach could be used in the pAIO system to silence the endogenous protein instead of using RNAi. It would improve the system in two ways: Firstly, the risk of off-target effects is further reduced. Secondly, genes that are difficult to knock down with RNAi become targetable.

7.4 Reduced levels of MLH1 R659Q due to decreased protein stability

When we used the pAIO system to replace endogenous MLH1 in Flp-In 293 T-REx cells with mutant GFP-MLH1 R659Q, we first of all observed significantly lowered protein levels compared to wt GFP-MLH1. Reduced GFP-MLH1 R659Q was found in whole cell extracts as well as in nuclear extracts. Similar results were obtained in previous studies. Takahashi and colleagues reported

expression levels of MLH1 R659Q of 25-75% after transient transfection in MLH-deficient HCT116 cells (12). Similarly, Hinrichsen and colleagues observed MLH1 R659Q levels of about 65% compared to wt when transiently expressed in MLH1-deficient 293T cells (95). Our data also suggest a decrease to 65%. Taken together, protein levels are clearly reduced due to the R659Q mutation. We wondered if this was caused by reduced transcription of the R659Q construct or by decreased mRNA stability. Because the mRNA levels were equal in all tested GFP-MLH1 variants, our results suggest that neither transcription nor mRNA levels are affected in the R659Q mutant. The same was observed in the study of Hinrichsen et al. (95). Although we cannot rule out that there is a translational block, it seems that the R659Q mutation affects the protein stability and might lead to increased protein turnover. Complex formation of proteins increases their stability and although the R695Q mutant still interacts strongly with PMS2, it showed impaired binding to FANCJ and FAN1. The loss of the ability to form those complexes might explain why this fraction undergoes early degradation, resulting in overall decreased levels of the R659Q mutant.

Interestingly, the K84E mutant showed increased protein levels in the cytoplasm, but not in the nucleus, which implies that the nuclear import of MLH1 molecules is tightly regulated. Obviously, increased levels in the cytoplasm do not automatically lead to increased levels in the nucleus. An explanation might be that strongly-increased MLH1 levels in the nucleus might impair MMR integrity, which has been observed in yeast (143).

7.5 Impaired MMR is not the cause of genome instability in MLH1 R659Q cells

When we initiated this study, our aim was to develop an improved pAIO system to study genotype-phenotype correlations. To test our system, we wanted to clarify the molecular pathogenesis of the MLH1 missense mutation R659Q. Our first results revealed decreased protein levels and we speculated that significantly decreased protein levels might also reduce MMR activity. To test this hypothesis, time-lapse mismatch repair assays with nuclear extracts of the different MLH1 variants were performed. Although the extracts containing the

R659Q mutant were slightly less active than those containing wt MLH1, the difference was not significant over three independent experiments. This suggests that, despite reduced protein levels of the R659Q mutant, it is MMR-proficient. MMR proficiency was also observed in other studies using the *in vitro* MMR assay. Hinrichsen et al. as well as Raevaara et al. observed normal MMR activity of the R659Q mutant compared to wt MLH1 (11,95). Takahashi and colleagues reported slightly decreased MMR activity of about 80% of the R659Q mutant (12). A question that remains is whether the *in vitro* MMR assay is a suitable tool to predict if slightly reduced MMR activities are still enough to prevent increased mutation rates *in vivo*. We tried to investigate this question by conducting a long-term experiment with the MLH1 variants. Over a period of four weeks, which accounts for about 40 cell divisions, we kept the cells in media with Dox to keep the pAIO system activated and repeatedly transfected the cells with siRNA to reduce endogenous MLH1 even further. Afterwards, several markers for MSI were analyzed. As we only observed a MSI phenotype in genomic DNA from 293T cells, which served as a positive control, but not in the MLH1 K84E mutant, which is known to cause MSI in patients, we came to the conclusion that this experimental setup is not suited to analyze the effect on genome stability of MLH1 mutations *in vivo*. This might have several causes. First, the time and number of cell divisions might not have been enough to cause MSI. Second, the remaining endogenous MLH1 might have been sufficient to prevent MSI because 10% of MLH1 are sufficient for the full MMR activity *in vitro* (72). To address this question, the generation of cell lines with mutated endogenous MLH1 by applying the CRISPR/Cas9 HSR approach might be helpful.

Taken together, the findings from this study and also the observations from previous studies strongly suggest that the R659Q mutation does not significantly impair MMR function and can therefore not be regarded as the driving force in tumor development in patients with the MLH1 R659Q mutation.

7.6 Loss of MLH1/FANCJ and FAN1 interactions might increase genomic instability

As the data from our study and others strongly imply that the MLH1 R659Q mutant is MMR-proficient, it remains unclear how carriers of this mutation acquire increased mutation rates that lead to an early onset of cancer. When we determined the interactome of the mutant, we found that binding to the DNA repair factors FANCJ and FAN1 was significantly impaired compared to GFP-MLH1 wt and K84E. FANCJ and FAN1 have been strongly associated with DNA interstrand crosslink (ICL) repair and it has been suggested that MLH1 plays a role here (59,62,65,124). When we tested the MLH1 variants against the ICL inducing agent MMC, we found that the K84E mutant and the R659Q mutant were more resistant than MLH1 wt. In the case of the R659Q mutant, we could confirm the result also in stably transfected 293T cells, which provide evidence in another cell system. We also observed a reduced arrest in G2/M phase after MMC treatment in the mutants, which was in line with increased cell survival rates. Together, these experiments support the idea of an altered DNA damage response. Reports from literature are very diverse when it comes to the question of how MLH1-depleted cells react towards MMC treatment. In some studies, MLH1-depleted cells were shown to be more resistant than wt cells, which confirms the findings of our study (51,52). However, there are also reports in the literature showing that MLH1-depleted cells are more sensitive to MMC treatment (50,162). Although all of these studies suggest that MLH1 plays a role in the repair of lesions that arise from MMC, the mechanism remains enigmatic. One model suggests that MLH1 plays a role in the recruitment of FANCJ to ICL lesions (57). It was observed that in MLH1-deficient HCT116 cells, the recruitment to laser-induced psoralen crosslinks was delayed compared to HCT116 + Chromosome 3 cells which have MLH1. Another model suggests that MLH1 plays a role in unhooking the lesion, but this model lacks experimental evidence (54). For both scenarios, MLH1 is directly involved in the repair of the ICL lesion and depletion of MLH1 might lead to sensitivity against ICL-inducing agents as was observed in some studies (50,162). Considering that other studies reported resistance against MMC treatment in MLH1-depleted cells, other

explanations of the mechanism seem plausible (51,52). A similar situation occurs when cells are treated with MNNG. MMR-deficient cells are much more resistant to MNNG than MMR-deficient cells. The attempted repair of the lesions creates conditions that are very toxic for the cell and lead to cell death, while unrepaired lesions are much less toxic, but at the expense of an increased mutation rate. It is tempting to argue that this might be the same for MMC-induced lesions. An attempt to repair those with the help of MLH1 might lead to the creation of more toxic DNA structures such as DNA DSBs than the repair process without MLH1. This might be mutagenic, but less toxic for the cells. It seems that an MLH1-dependent repair process of MMC induced lesions requires ATPase active MLH1 as well as the capability to interact with FANCJ and FAN1, because both mutations altered the DDR after MMC treatment. With the data collected so far it is difficult to determine if only the interaction with FANCJ, FAN1 or both is required for the MLH1-dependent repair of the lesions. To answer this question, an MLH1 mutant that can only interact with FAN1 or FANCJ is required. Xie and colleagues reported that the MLH1 mutant L607H, which they found to be deficient in binding to FANCJ, was more sensitive to MMC treatment than wt MLH1 in HCT116 cells. In the same assay, however, MLH1-deficient cells were even more resistant (163). Unfortunately, there is no information about the interaction with FAN1. It appears that the genetic background of HCT116 cells significantly influences the outcome of cell survival assays with MMC depending on the MLH1 status, because the studies that showed resistance after MLH1 depletion and also ours with different mutants were conducted in several cell lines such as 293, HeLa and U2OS. Another explanation for MMC resistance in MLH1-depleted or MLH1 mutant cells might be that the interaction of MLH1 inhibits the activity of FANCJ and FAN1. A slower but controlled repair of the ICL lesions might be important for accurate repair and could prevent the creation of toxic lesions. A summary of the different models is given in Figure 30.

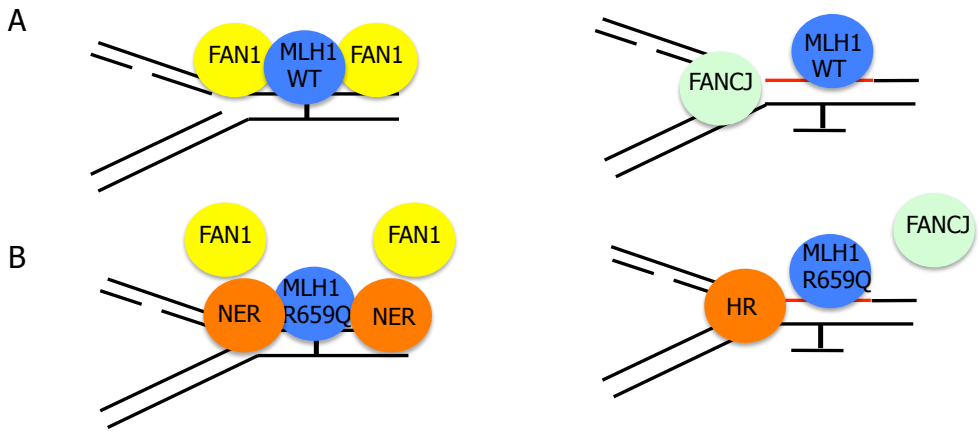


Figure 30: Roles of MLH1 variants in ICL repair

(**A**) After FA core complex activation, MLH1 recognizes the ICL lesion and recruits FAN1 to unhook the lesion followed by TLS-dependent repair. The resulting structures are resolved by FANCJ, which is also recruited by MLH1. (**B**) MLH1 R659Q cannot recruit FAN1 and the lesion is left unaddressed or canonical ICL repair factors take over. Later on FANCJ is not recruited for resolution and remodeling of the replication fork.

Also, the initiation of DNA damage signaling might play a role in the resistance of the MLH1 R659Q mutant. If it cannot bind to FAN1 and FANCJ, MLH1 might lose the interaction with the lesion and, in turn, signaling pathways may not be triggered.

Although it seems clear that a lack of MLH1 or mutations that affect its activity or its interaction with FANCJ and FAN1 alter the DDR towards lesions induced by MMC, this mechanism clearly requires more investigation. Still, the data provide a potential explanation of how carriers of the MLH1 R659Q variant acquire mutations that trigger tumor development. ICL lesions induced by endogenous or exogenous agents such as acetaldehyde after alcohol consumption might be more repaired in a way that ensures cell survival, but at the expense of increased mutation rates (164).

7.7 AP-MS of MSH2-eGFP identifies WDHD1 as an interactor of MutSα

In order to identify new factors that might facilitate MMR *in vivo* and be of importance for non-canonical MMR, we fused the MMR factors MLH1 and MSH2 to GFP or eGFP, respectively. After pulldown with GFP-Trap followed by LC-MS, we could identify many known interaction partners, but also found some new interesting candidates. One of them is WDHD1. Follow-up experiments showed that MSH2 knockdown reduces WDHD1 levels. Moreover, WDHD1 levels decrease after MNNG treatment. Work of Chen and colleagues also demonstrated the interaction of WDHD1 with MSH2. They could map the interaction domain to the C-terminal region of WDHD1 and in MSH2 to the region between the clamp domain and the ATPase domain. Moreover, they observed resistance of WDHD1-depleted cells towards 6-TG similar to MSH2 depletion (148). Together, these results suggest that WDHD1 plays a role in *in vivo* MMR. However, the resistance of WDHD1 could also be MMR-independent. Further studies are needed to explore a potential role of WDHD1 in MMR *in vivo*. First of all, it would be interesting to see how WDHD1 depletion affects cell survival after MNNG treatment. If WDHD1 is indeed needed for *in vivo* MMR, depletion should lead to resistance. Surprisingly, MNNG treatment resulted in decreased levels of WDHD1. Chen and colleagues also reported that WDHD1 is mutated in numerous cancers and might contribute to tumorigenesis. It might be of great interest to test if those cancers show MSI instability, which could be another hint towards the involvement of WDHD1 in MMR.

We also found the helicases MCM8 and MCM9 in our AP-MS data for MSH2. Researchers have for a long time searched for helicases involved in MMR in eukaryotes that might be functional homologous to the UvrD helicase in bacterial MMR. Interestingly, a study by Traver et al. now suggests that MCM9 helicase is needed for *in vivo* MMR in a way that its helicase activity unwinds the mismatch-containing DNA region, thus catalyzing the removal of the mismatch. They showed that MCM9 strongly interacts with MMR initiation factors and provide data that MCM9 is needed to recruit MLH1 to chromatin (31).

7.8 AP-MS does not identify DNA damage-dependent change of MMR interactomes

In addition to gathering information about the general interactome of MLH1 and MSH2, we treated the cells with MNNG to track changes and to identify factors important for non-canonical MMR. However, when comparing the data of untreated and treated cells, we could not find significant changes. This might have several reasons. First of all, if changes occur, they might only be transient or weak and thus might be lost during the pulldown. Furthermore, there might only be minor changes that fall below the threshold and are filtered out. The strong binding of MLH1 to PMS2 and PMS1 or of MSH2 to MSH3 and MSH6 might mask the appearance or loss of factors after MNNG treatment. Different approaches to study transient and weak interactions have been developed including the use of crosslinking agents or proximity-dependent labeling techniques such as BioID or APEX.

Prominent protein-protein crosslinkers include the imidoester crosslinker dimethyl suberimidate or the N-hydroxysuccinimide-ester crosslinker BS3 and formaldehyde. Each of these crosslinkers reacts with the amino groups of lysines to form covalent bonds. Another method of *in vivo* crosslinking applies photo-reactive amino acid analogs. Cells are grown with photoreactive diazirine analogs to methionine as well as leucine and are incorporated into the proteins during translation. Exposure to UV light activates the diazirines, which then bind to interacting proteins (165). In this study, we explored BioID and Apex as methods to label proteins in close proximity, which can be later pulled down by affinity purification.

7.9 Drawbacks of BioID and APEX technologies

As we could not identify changes in the interactome of MLH1 and MSH2 following MNNG treatment, we decided to explore BioID and APEX for our purposes (128,129). First, we found that BirA* fused N-terminally to MLH1 was inactive and moreover impaired the stability of MLH1 and its localization. There might be ways to circumvent these problems. The usage of a longer linker between the two proteins could help them fold more independently and thus

help sustain their functions. In general, BirA* could also be fused to the C-terminus which might solve the problem. In the case of MLH1, this is not recommended, because the last residue of yeast MLH1 was shown to contribute to the endonuclease function of PMS1 (90). Also, studies with human MLH1 have demonstrated that C-terminal tagging of MLH1 impairs its function (142). Alternatively, one could fuse BirA* within the MLH1 sequence between the functional domains. For instance, the mid-region between the binding site for MutS homologs and EXO1 appears suitable for such an approach. However, experience from other labs suggests that these changes might not be successful to obtain an active and functional BirA*-MLH1 fusion protein (Richard Lutz and Kerstin Gari, IMCR University of Zurich, personal communication). Hence, it might be more promising to use PMS2-BirA* which we have shown to be active. As a next step, it needs to be verified whether PMS2-BirA* is functional in MMR and localizes to the nucleus.

As another approach based on proximity labeling, we explored the APEX system. In contrast to BirA*-MLH1, APEX2-MLH1 was active and behaved as wt MLH1. We used the system to study changes in the interactome of MLH1 after MNNG and MMC treatment and could identify some interesting factors. However, when we introduced an APEX2-GFP as another control, we found that APEX unspecifically labeled nearly every protein we tested. One could argue that a Western Blot might not be enough to conclude that APEX2 is too unspecific, but LC-MS data from colleagues (Kerstin Gari and Richard Lutz, IMCR University of Zurich) who also used our APEX2-pAIO plasmid for protein-protein interaction studies confirmed that APEX2 is too unspecific and seems to label every protein (personal communication). Instead of APEX2-GFP, they used APEX2 alone but also with that approach, very well described binding proteins of the bait proteins were found in the APEX2 control. With these findings, we concluded that the background in this approach is too strong to search for important new interactions after DNA damage and we thought about ways to limit the unspecific activity. One idea is to fuse APEX2 to another DNA repair factor such as TDG. This might help to limit the labeling of proteins and to restrict the area. Another possibility is to use APEX instead of APEX2. APEX2 has been developed to increase its activity (132). As in our case the strong activity of APEX2 could be

the reason for its low specificity, a limited generation of biotinphenol radicals by using less active APEX could be a solution.

7.10 Alternative strategies to study the MMR interactome after DNA damage

Although using PMS2-BirA* and APEX instead of APEX2 to reduce its activity provide potential solutions to the drawbacks, it is worthwhile to think about alternative approaches to study the MMR interactome in different aspects of DNA metabolism. Previous work from our group has demonstrated that non-canonical MMR cannot only be activated by MNNG treatment of living cells, but also by plasmids containing U-G mismatches if BER is inactivated (74). Using this approach, it might be possible to induce non-canonical MMR in nuclear extracts and to pull down important factors involved in the process either by targeting the plasmid or tagged proteins such as GFP-MLH1 or MSH2-eGFP. If the system proves to be informative, it might also be interesting to see what other proteins take part in canonical MMR using nuclear extracts.

A very elegant method to study the proteins involved in ICL repair called CHROMASS (chromatin mass spectrometry) has been recently described by Raschle and colleagues (166). They used replicating Xenopus egg extracts and ICL-containing chromatin and analyzed the proteins that bound to the chromatin and studied how they might contribute to ICL repair. It is not difficult to imagine that this system could be adapted to study the proteome of non-canonical MMR. Indeed, our group had used Xenopus egg extracts before to study the hypothesis of futile cycling in a replicative system (44).

8 Conclusion and Outlook

The pAIO system has proven to be a very useful technique to study the phenotype of gene variants. With the help of the first generation pAIO, our lab could study polymerase δ variants in a human system *in vivo* and the second generation pAIO plasmid helped to study the MLH1 R659Q mutation that had been contradictive for many years. We were able to integrate some very useful features into the new pAIO plasmid, which make the system much easier to work with and more efficient. If an entry clone plasmid is available, it is possible to clone the desired plasmids within a week and stable cell lines are available within three to four weeks. This is very fast compared to other techniques used to replace or reconstitute gene variants. Although the cDNA in our system is driven by an exogenous CMV promoter and lacks endogenous regulation, we found ways to adapt the expression level of our gene of interest to be comparable to that of the endogenous protein. We used our cell system for various cell biology and biochemistry assays and it is fair to say that we could use the system for nearly every experiment that is based on or uses cell lines.

Although the protein replacement based on RNAi worked very efficiently for polymerase δ as well as for MLH1 when supported with siRNA, some proteins are difficult to knock down with RNAi and there is always the risk of off-target effects. Hence, it might be worthwhile to think about another improvement of our system by combining it with the latest development of the CRISPR/Cas9 technique. It should be possible to generate a cell line that expresses a nuclease-dead Cas9 fusion protein that induces transcriptional repression through epigenetic silencing such as KRAB. Now, instead of cloning an shRNA into the pAIO vector, an sgRNA guides the fusion protein to the promoter of the gene of interest, which leads to its repression, while a variant of the gene is expressed driven by the CMV promoter in the pAIO plasmid.

We have used the second generation pAIO system to produce stable cell lines in which, upon Dox induction endogenous MLH1 is replaced by GFP-tagged

variants. One of the variants we created is the MLH1 R659Q mutant. The mutation has been found in several HNPCC families but the molecular pathology remained unclear. With the help of our system, we could confirm that the mutation results in reduced levels of MLH1 in whole cell extracts as well as in the nucleus, but nevertheless it is MMR-active. A very striking difference was observed when we studied its interactions with the DNA repair factors FANCJ and FAN1. Compared to wt GFP-MLH1, binding to both was strongly reduced. As both factors are associated with ICL repair, we checked for the response towards ICL-inducing MMC treatment and found resistance in cell survival assays as well as a reduced G2/M phase arrest. It appears that the limited binding to either FANCJ and/or FAN1 alters the MLH1-dependent DDR towards MMC induced lesions. This might give a hint as to how R659Q mutation carriers acquire mutations despite the fact that the cells are MMR-proficient. In further experiments, we should try to decipher the molecular mechanism of how the MLH1 R659Q mutant affects ICL lesion processing. This understanding might then help to develop strategies on how to prevent or decrease mutations in mutation carriers. At the moment, it seems that a reduction of ICLs could reduce the number of mutations. Although it is difficult to prevent ICLs that arise from endogenous sources, it might be worthwhile to reduce exogenous sources of ICLs such as acetaldehyde, which is the degradation product of alcohol consumption.

In the second part of the study we focused on the MMR interactome and were especially interested in the changes of the interactome after induction of non-canonical MMR causing DNA damage. First, we used GFP- and eGFP-tagged MMR proteins and enriched those and their binding partners by pulldown with GFP-Trap. Although we could identify some intersting candidates such as WDHD1 or MCM9, which have later been proposed to play a role in *in vivo* MMR (31, 139), this approach did not help to identify changes in the MMR interactome after DNA damage. Hence, other techniques were needed and we explored the BioID as well as the APEX technique for our purposes. We could show that both approaches can be applied to biotinylate endogenous polypeptides and managed to enrich these labeled proteins. However, both methods had their drawbacks and still need to be improved before they can be used reliably on a regular basis. It seems

that the APEX method in particular can only be applied if targeting a specific compartment of the cells as in the case of the mitochondrial proteome (129). Only if the protein of interest focuses on one or more very defined sites can the unspecific binding be reduced and only the proteins that localize to those sites become labeled and will be identified in later analyses. In order to study the interactome of MMR after DNA damage and thus to better understand non-canonical MMR, it will be necessary to explore other techniques such as protein-protein crosslinking or CHROMASS (157).

9 References

1. Weinberg, R. A. (2007) *The biology of cancer*, Garland Science, New York
2. Alberts, B., Wilson, J. H., and Hunt, T. (2008) *Molecular biology of the cell*, 5th ed., Garland Science, New York
3. Banks, K. C., Moline, J. J., Marvin, M. L., Newlin, A. C., and Vogel, K. J. (2013) *Fam Cancer* **12**, 1-18
4. Knudson, A. G., Jr. (1971) *Proc Natl Acad Sci U S A* **68**, 820-823
5. Shlien, A., Campbell, B. B., de Borja, R., Alexandrov, L. B., Merico, D., Wedge, D., Van Loo, P., Tarpey, P. S., Coupland, P., Behjati, S., Pollett, A., Lipman, T., Heidari, A., Deshmukh, S., Avitzur, N., Meier, B., Gerstung, M., Hong, Y., Merino, D. M., Ramakrishna, M., Remke, M., Arnold, R., Panigrahi, G. B., Thakkar, N. P., Hodel, K. P., Henninger, E. E., Goksenin, A. Y., Bakry, D., Charames, G. S., Druker, H., Lerner-Ellis, J., Mistry, M., Dvir, R., Grant, R., Elhasid, R., Farah, R., Taylor, G. P., Nathan, P. C., Alexander, S., Ben-Shachar, S., Ling, S. C., Gallinger, S., Constantini, S., Dirks, P., Huang, A., Scherer, S. W., Grundy, R. G., Durno, C., Aronson, M., Gartner, A., Meyn, M. S., Taylor, M. D., Pursell, Z. F., Pearson, C. E., Malkin, D., Futreal, P. A., Stratton, M. R., Bouffet, E., Hawkins, C., Campbell, P. J., and Tabori, U. (2015) *Nat Genet* **47**, 257-262
6. Kuismanen, S. A., Holmberg, M. T., Salovaara, R., de la Chapelle, A., and Peltomaki, P. (2000) *The American journal of pathology* **156**, 1773-1779
7. Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Krieger, M., and Scott, M. P. (2004) *Molecular Cell Biology*, 5th ed., W.H. Freeman and Company, New York
8. Carethers, J. M., and Stoffel, E. M. (2015) *World journal of gastroenterology : WJG* **21**, 9253-9261
9. Howlader, N., Noone, A., Krapcho, M., Garshell, J., Miller, D., Altekruse, S., Kosary, C., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D., Chen, H., Feuer, E., and Cronin, K. (1975-2012) SEER cancer statistics review. in *NIH publication*, U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute, Bethesda, Md.
10. Prat, J., Ribe, A., and Gallardo, A. (2005) *Human pathology* **36**, 861-870
11. Raevaara, T. E., Korhonen, M. K., Lohi, H., Hampel, H., Lynch, E., Lonnqvist, K. E., Holinski-Feder, E., Sutter, C., McKinnon, W., Duraisamy, S., Gerdes, A. M., Peltomaki, P., Kohonen-Ccorish, M., Mangold, E., Macrae, F., Greenblatt, M., de la Chapelle, A., and Nystrom, M. (2005) *Gastroenterology* **129**, 537-549
12. Takahashi, M., Shimodaira, H., Andreutti-Zaugg, C., Iggo, R., Kolodner, R. D., and Ishioka, C. (2007) *Cancer Res* **67**, 4595-4604
13. Jiricny, J. (2006) *Nature reviews. Molecular cell biology* **7**, 335-346
14. Vasen, H. F., Blanco, I., Aktan-Collan, K., Gopie, J. P., Alonso, A., Aretz, S., Bernstein, I., Bertario, L., Burn, J., Capella, G., Colas, C., Engel, C., Frayling, I. M., Genuardi, M., Heinimann, K., Hes, F. J., Hodgson, S. V., Karagiannis, J. A., Lalloo, F., Lindblom, A., Mecklin, J. P., Moller, P., Myrhoj, T., Nagengast, F. M., Parc, Y., Ponz de Leon, M., Renkonen-Sinisalo, L., Sampson, J. R.,

Stormorken, A., Sijmons, R. H., Tejpar, S., Thomas, H. J., Rahner, N., Wijnen, J. T., Jarvinen, H. J., and Moslein, G. (2013) *Gut* **62**, 812-823
15. Peltomaki, P., and Vasen, H. (2004) *Disease markers* **20**, 269-276
16. http://www.insight-group.org.
17. Vasen, H. F., Mecklin, J. P., Khan, P. M., and Lynch, H. T. (1991) *Diseases of the colon and rectum* **34**, 424-425
18. Vasen, H. F., Watson, P., Mecklin, J. P., and Lynch, H. T. (1999) *Gastroenterology* **116**, 1453-1456
19. Pinol, V., Castells, A., Andreu, M., Castellvi-Bel, S., Alenda, C., Llor, X., Xicola, R. M., Rodriguez-Moranta, F., Paya, A., Jover, R., and Bessa, X. (2005) *JAMA* **293**, 1986-1994
20. Ribic, C. M., Sargent, D. J., Moore, M. J., Thibodeau, S. N., French, A. J., Goldberg, R. M., Hamilton, S. R., Laurent-Puig, P., Gryfe, R., Shepherd, L. E., Tu, D., Redston, M., and Gallinger, S. (2003) *N Engl J Med* **349**, 247-257
21. Kunkel, T. A. (2004) *J Biol Chem* **279**, 16895-16898
22. Levinson, G., and Gutman, G. A. (1987) *Molecular biology and evolution* **4**, 203-221
23. Iyer, R. R., Pluciennik, A., Burdett, V., and Modrich, P. L. (2006) *Chemical reviews* **106**, 302-323
24. Lahue, R. S., Au, K. G., and Modrich, P. (1989) *Science* **245**, 160-164
25. Kunkel, T. A., and Erie, D. A. (2005) *Annu Rev Biochem* **74**, 681-710
26. Kadyrov, F. A., Dzantiev, L., Constantin, N., and Modrich, P. (2006) *Cell* **126**, 297-308
27. Ghodgaonkar, M. M., Lazzaro, F., Olivera-Pimentel, M., Artola-Boran, M., Cejka, P., Reijns, M. A., Jackson, A. P., Plevani, P., Muzi-Falconi, M., and Jiricny, J. (2013) *Mol Cell* **50**, 323-332
28. Lujan, S. A., Williams, J. S., Clausen, A. R., Clark, A. B., and Kunkel, T. A. (2013) *Mol Cell* **50**, 437-443
29. Repmann, S., Olivera-Harris, M., and Jiricny, J. (2015) *J Biol Chem* **290**, 9986-9999
30. Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G. M. (2013) *Cell* **153**, 590-600
31. Traver, S., Coulombe, P., Peiffer, I., Hutchins, J. R., Kitzmann, M., Latreille, D., and Mechali, M. (2015) *Mol Cell* **59**, 831-839
32. Lopez-Contreras, A. J., Ruppen, I., Nieto-Soler, M., Murga, M., Rodriguez-Acebes, S., Remeseiro, S., Rodrigo-Perez, S., Rojas, A. M., Mendez, J., Munoz, J., and Fernandez-Capetillo, O. (2013) *Cell Rep* **3**, 1105-1116
33. Schopf, B., Bregenhorn, S., Quivy, J. P., Kadyrov, F. A., Almouzni, G., and Jiricny, J. (2012) *Proc Natl Acad Sci U S A* **109**, 1895-1900
34. Zhang, M., Xiang, S., Joo, H. Y., Wang, L., Williams, K. A., Liu, W., Hu, C., Tong, D., Haakenson, J., Wang, C., Zhang, S., Pavlovicz, R. E., Jones, A., Schmidt, K. H., Tang, J., Dong, H., Shan, B., Fang, B., Radhakrishnan, R., Glazer, P. M., Matthias, P., Koomen, J., Seto, E., Bepler, G., Nicosia, S. V., Chen, J., Li, C., Gu, L., Li, G. M., Bai, W., Wang, H., and Zhang, X. (2014) *Mol Cell* **55**, 31-46
35. Valeri, N., Gasparini, P., Braconi, C., Paone, A., Lovat, F., Fabbri, M., Sumani, K. M., Alder, H., Amadori, D., Patel, T., Nuovo, G. J., Fishel, R., and Croce, C. M. (2010) *Proc Natl Acad Sci U S A* **107**, 21098-21103

36. Valeri, N., Gasparini, P., Fabbri, M., Braconi, C., Veronese, A., Lovat, F., Adair, B., Vannini, I., Fanini, F., Bottoni, A., Costinean, S., Sandhu, S. K., Nuovo, G. J., Alder, H., Gafa, R., Calore, F., Ferracin, M., Lanza, G., Volinia, S., Negrini, M., McIlhatton, M. A., Amadori, D., Fishel, R., and Croce, C. M. (2010) *Proc Natl Acad Sci U S A* **107**, 6982-6987
37. Duckett, D. R., Drummond, J. T., Murchie, A. I., Reardon, J. T., Sancar, A., Lilley, D. M., and Modrich, P. (1996) *Proc Natl Acad Sci U S A* **93**, 6443-6447
38. Zlatanou, A., Despras, E., Braz-Petta, T., Boubakour-Azzouz, I., Pouvelle, C., Stewart, G. S., Nakajima, S., Yasui, A., Ishchenko, A. A., and Kannouche, P. L. (2011) *Mol Cell* **43**, 649-662
39. Stojic, L., Brun, R., and Jiricny, J. (2004) *DNA Repair (Amst)* **3**, 1091-1101
40. Kondo, N., Takahashi, A., Ono, K., and Ohnishi, T. (2010) *J Nucleic Acids* **2010**, 543531
41. Kaina, B., Christmann, M., Naumann, S., and Roos, W. P. (2007) *DNA Repair (Amst)* **6**, 1079-1099
42. Stojic, L., Cejka, P., and Jiricny, J. (2005) *Cell Cycle* **4**, 473-477
43. Stojic, L., Mojas, N., Cejka, P., Di Pietro, M., Ferrari, S., Marra, G., and Jiricny, J. (2004) *Genes Dev* **18**, 1331-1344
44. Olivera Harris, M., Kallenberger, L., Artola Boran, M., Enoiu, M., Costanzo, V., and Jiricny, J. (2015) *DNA Repair (Amst)* **28**, 1-7
45. Yoshioka, K., Yoshioka, Y., and Hsieh, P. (2006) *Mol Cell* **22**, 501-510
46. Fink, D., Nebel, S., Aebi, S., Zheng, H., Cenni, B., Nehme, A., Christen, R. D., and Howell, S. B. (1996) *Cancer Res* **56**, 4881-4886
47. Papouli, E., Cejka, P., and Jiricny, J. (2004) *Cancer Res* **64**, 3391-3394
48. Wang, D., and Lippard, S. J. (2005) *Nat Rev Drug Discov* **4**, 307-320
49. Fiumicino, S., Martinelli, S., Colussi, C., Aquilina, G., Leonetti, C., Crescenzi, M., and Bignami, M. (2000) *Int J Cancer* **85**, 590-596
50. Williams, S. A., Wilson, J. B., Clark, A. P., Mitson-Salazar, A., Tomashevski, A., Ananth, S., Glazer, P. M., Semmes, O. J., Bale, A. E., Jones, N. J., and Kupfer, G. M. (2011) *Hum Mol Genet* **20**, 4395-4410
51. Huang, M., Kennedy, R., Ali, A. M., Moreau, L. A., Meetei, A. R., D'Andrea, A. D., and Chen, C. C. (2011) *DNA Repair (Amst)* **10**, 1203-1212
52. Peng, M., Xie, J., Ucher, A., Stavnezer, J., and Cantor, S. B. (2014) *Embo J* **33**, 1698-1712
53. Moldovan, G. L., and D'Andrea, A. D. (2009) *Annu Rev Genet* **43**, 223-249
54. Bergstralh, D. T., and Sekelsky, J. (2008) *Trends Genet* **24**, 70-76
55. Clauson, C., Scharer, O. D., and Niedernhofer, L. (2013) *Cold Spring Harb Perspect Biol* **5**, a012732
56. Muniandy, P. A., Liu, J., Majumdar, A., Liu, S. T., and Seidman, M. M. (2010) *Crit Rev Biochem Mol Biol* **45**, 23-49
57. Suhasini, A. N., Sommers, J. A., Muniandy, P. A., Coulombe, Y., Cantor, S. B., Masson, J. Y., Seidman, M. M., and Brosh, R. M., Jr. (2013) *Mol Cell Biol* **33**, 2212-2227
58. Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R., and Jiricny, J. (2007) *J Biol Chem* **282**, 2976-2986
59. Peng, M., Litman, R., Xie, J., Sharma, S., Brosh, R. M., Jr., and Cantor, S. B. (2007) *Embo J* **26**, 3238-3249

60. Pizzolato, J., Mukherjee, S., Scharer, O. D., and Jiricny, J. (2015) *J Biol Chem* **290**, 22602-22611
61. Wang, R., Persky, N. S., Yoo, B., Ouerfelli, O., Smogorzewska, A., Elledge, S. J., and Pavletich, N. P. (2014) *Science* **346**, 1127-1130
62. Cantor, S. B., Bell, D. W., Ganesan, S., Kass, E. M., Drapkin, R., Grossman, S., Wahrer, D. C., Sgroi, D. C., Lane, W. S., Haber, D. A., and Livingston, D. M. (2001) *Cell* **105**, 149-160
63. Levitus, M., Waisfisz, Q., Godthelp, B. C., de Vries, Y., Hussain, S., Wiegant, W. W., Elghalbzouri-Maghrani, E., Steltenpool, J., Rooimans, M. A., Pals, G., Arwert, F., Mathew, C. G., Zdzienicka, M. Z., Hiom, K., De Winter, J. P., and Joenje, H. (2005) *Nat Genet* **37**, 934-935
64. Yoshikiyo, K., Kratz, K., Hirota, K., Nishihara, K., Takata, M., Kurumizaka, H., Horimoto, S., Takeda, S., and Jiricny, J. (2010) *Proc Natl Acad Sci U S A* **107**, 21553-21557
65. Kratz, K., Schopf, B., Kaden, S., Sendoel, A., Eberhard, R., Lademann, C., Cannavo, E., Sartori, A. A., Hengartner, M. O., and Jiricny, J. (2010) *Cell* **142**, 77-88
66. Zhang, H., Richards, B., Wilson, T., Lloyd, M., Cranston, A., Thorburn, A., Fishel, R., and Meuth, M. (1999) *Cancer Res* **59**, 3021-3027
67. Duckett, D. R., Bronstein, S. M., Taya, Y., and Modrich, P. (1999) *Proc Natl Acad Sci U S A* **96**, 12384-12388
68. Peters, A. C., Young, L. C., Maeda, T., Tron, V. A., and Andrew, S. E. (2003) *DNA Repair (Amst)* **2**, 427-435
69. Hickman, M. J., and Samson, L. D. (1999) *Proc Natl Acad Sci U S A* **96**, 10764-10769
70. Gong, J. G., Costanzo, A., Yang, H. Q., Melino, G., Kaelin, W. G., Jr., Levrero, M., and Wang, J. Y. (1999) *Nature* **399**, 806-809
71. Shimodaira, H., Yoshioka-Yamashita, A., Kolodner, R. D., and Wang, J. Y. (2003) *Proc Natl Acad Sci U S A* **100**, 2420-2425
72. Cejka, P., Stojic, L., Mojas, N., Russell, A. M., Heinimann, K., Cannavo, E., di Pietro, M., Marra, G., and Jiricny, J. (2003) *Embo J* **22**, 2245-2254
73. Schroering, A. G., and Williams, K. J. (2008) *DNA Repair (Amst)* **7**, 951-969
74. Pena-Diaz, J., Bregenhorn, S., Ghodgaonkar, M., Follonier, C., Artola-Boran, M., Castor, D., Lopes, M., Sartori, A. A., and Jiricny, J. (2012) *Mol Cell* **47**, 669-680
75. Bregenhorn, S., Kallenberger, L., Artola-Boran, M., Pena-Diaz, J., and Jiricny, J. (2016) *Nucleic Acids Res*
76. Aaltonen, L. A., Peltomaki, P., Mecklin, J. P., Jarvinen, H., Jass, J. R., Green, J. S., Lynch, H. T., Watson, P., Tallqvist, G., Juhola, M., and et al. (1994) *Cancer Res* **54**, 1645-1648
77. Vogelstein, B., and Kinzler, K. W. (1993) *Trends Genet* **9**, 138-141
78. Malkhosyan, S., Rampino, N., Yamamoto, H., and Perucho, M. (1996) *Nature* **382**, 499-500
79. Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., Vogelstein, B., and et al. (1995) *Science* **268**, 1336-1338
80. Peltomaki, P. (2001) *Mutat Res* **488**, 77-85
81. Narayan, S., and Roy, D. (2003) *Mol Cancer* **2**, 41
82. Andrew, S. (1999) *Clin Genet* **56**, 186-188

83. Wei, K., Kucherlapati, R., and Edelmann, W. (2002) *Trends Mol Med* **8**, 346-353
84. Lee, K., Tosti, E., and Edelmann, W. (2015) *DNA Repair (Amst)*
85. Han, H. J., Maruyama, M., Baba, S., Park, J. G., and Nakamura, Y. (1995) *Hum Mol Genet* **4**, 237-242
86. Olipitz, W., and Sill, H. (2000) *Hum Mol Genet* **9**, 321
87. Kondo, E., Horii, A., and Fukushige, S. (2001) *Nucleic Acids Res* **29**, 1695-1702
88. Raschle, M., Dufner, P., Marra, G., and Jiricny, J. (2002) *J Biol Chem* **277**, 21810-21820
89. Tran, P. T., and Liskay, R. M. (2000) *Mol Cell Biol* **20**, 6390-6398
90. Gueneau, E., Dherin, C., Legrand, P., Tellier-Lebegue, C., Gilquin, B., Bonnesoeur, P., Londino, F., Quemener, C., Le Du, M. H., Marquez, J. A., Moutiez, M., Gondry, M., Boiteux, S., and Charbonnier, J. B. (2013) *Nat Struct Mol Biol* **20**, 461-468
91. Lynch, H. T., and de la Chapelle, A. (2003) *N Engl J Med* **348**, 919-932
92. Shcherbakova, P. V., and Kunkel, T. A. (1999) *Mol Cell Biol* **19**, 3177-3183
93. Shimodaira, H., Filosi, N., Shibata, H., Suzuki, T., Radice, P., Kanamaru, R., Friend, S. H., Kolodner, R. D., and Ishioka, C. (1998) *Nat Genet* **19**, 384-389
94. Wahlberg, S. S., Schmeits, J., Thomas, G., Loda, M., Garber, J., Syngal, S., Kolodner, R. D., and Fox, E. (2002) *Cancer Res* **62**, 3485-3492
95. Hinrichsen, I., Brieger, A., Trojan, J., Zeuzem, S., Nilbert, M., and Plotz, G. (2013) *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**, 2432-2441
96. Muller, W., Burgart, L. J., Krause-Paulus, R., Thibodeau, S. N., Almeida, M., Edmonston, T. B., Boland, C. R., Sutter, C., Jass, J. R., Lindblom, A., Lubinski, J., MacDermot, K., Sanders, D. S., Morreau, H., Muller, A., Oliani, C., Orntoft, T., Ponz De Leon, M., Rosty, C., Rodriguez-Bigas, M., Ruschoff, J., Ruszkiewicz, A., Sabourin, J., Salovaara, R., and Moslein, G. (2001) *Fam Cancer* **1**, 87-92
97. de La Chapelle, A. (2002) *J Clin Oncol* **20**, 897-899
98. Ellison, A. R., Lofing, J., and Bitter, G. A. (2001) *Hum Mol Genet* **10**, 1889-1900
99. Guerrette, S., Wilson, T., Gradia, S., and Fishel, R. (1998) *Mol Cell Biol* **18**, 6616-6623
100. Guerrette, S., Acharya, S., and Fishel, R. (1999) *J Biol Chem* **274**, 6336-6341
101. Nystrom-Lahti, M., Perrera, C., Raschle, M., Panyushkina-Seiler, E., Marra, G., Curci, A., Quaresima, B., Costanzo, F., D'Urso, M., Venuta, S., and Jiricny, J. (2002) *Genes Chromosomes Cancer* **33**, 160-167
102. Kariola, R., Otway, R., Lonnqvist, K. E., Raevaara, T. E., Macrae, F., Vos, Y. J., Kohonen-Corish, M., Hofstra, R. M., and Nystrom-Lahti, M. (2003) *Hum Genet* **112**, 105-109
103. Kariola, R., Raevaara, T. E., Lonnqvist, K. E., and Nystrom-Lahti, M. (2002) *Hum Mol Genet* **11**, 1303-1310
104. Kondo, E., Suzuki, H., Horii, A., and Fukushige, S. (2003) *Cancer Res* **63**, 3302-3308
105. Schmutte, C., Sadoff, M. M., Shim, K. S., Acharya, S., and Fishel, R. (2001) *J Biol Chem* **276**, 33011-33018

106. Holmes, J., Jr., Clark, S., and Modrich, P. (1990) *Proc Natl Acad Sci U S A* **87**, 5837-5841
107. Thomas, D. C., Roberts, J. D., and Kunkel, T. A. (1991) *J Biol Chem* **266**, 3744-3751
108. Li, G. M., and Modrich, P. (1995) *Proc Natl Acad Sci U S A* **92**, 1950-1954
109. Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P., and Jiricny, J. (1999) *J Biol Chem* **274**, 32368-32375
110. Iaccarino, I., Palombo, F., Drummond, J., Totty, N. F., Hsuan, J. J., Modrich, P., and Jiricny, J. (1996) *Curr Biol* **6**, 484-486
111. Brieger, A., Adryan, B., Wolpert, F., Passmann, S., Zeuzem, S., and Trojan, J. (2010) *Proteomics* **10**, 3343-3355
112. Fields, S., and Song, O. (1989) *Nature* **340**, 245-246
113. Joung, J. K., Ramm, E. I., and Pabo, C. O. (2000) *Proc Natl Acad Sci U S A* **97**, 7382-7387
114. Gietz, R. D., Triggs-Raine, B., Robbins, A., Graham, K. C., and Woods, R. A. (1997) *Mol Cell Biochem* **172**, 67-79
115. Wang, J., Huo, K., Ma, L., Tang, L., Li, D., Huang, X., Yuan, Y., Li, C., Wang, W., Guan, W., Chen, H., Jin, C., Wei, J., Zhang, W., Yang, Y., Liu, Q., Zhou, Y., Zhang, C., Wu, Z., Xu, W., Zhang, Y., Liu, T., Yu, D., Chen, L., Zhu, D., Zhong, X., Kang, L., Gan, X., Yu, X., Ma, Q., Yan, J., Zhou, L., Liu, Z., Zhu, Y., Zhou, T., He, F., and Yang, X. (2011) *Mol Syst Biol* **7**, 536
116. Mac Partlin, M., Homer, E., Robinson, H., McCormick, C. J., Crouch, D. H., Durant, S. T., Matheson, E. C., Hall, A. G., Gillespie, D. A., and Brown, R. (2003) *Oncogene* **22**, 819-825
117. Rolland, T., Tasan, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., Kamburov, A., Ghiassian, S. D., Yang, X., Ghamsari, L., Balcha, D., Begg, B. E., Braun, P., Brehme, M., Broly, M. P., Carvunis, A. R., Convery-Zupan, D., Corominas, R., Coulombe-Huntington, J., Dann, E., Dreze, M., Dricot, A., Fan, C., Franzosa, E., Gebreab, F., Gutierrez, B. J., Hardy, M. F., Jin, M., Kang, S., Kiros, R., Lin, G. N., Luck, K., MacWilliams, A., Menche, J., Murray, R. R., Palagi, A., Poulin, M. M., Rambout, X., Rasla, J., Reichert, P., Romero, V., Ruyssinck, E., Sahalie, J. M., Scholz, A., Shah, A. A., Sharma, A., Shen, Y., Spirohn, K., Tam, S., Tejeda, A. O., Trigg, S. A., Twizere, J. C., Vega, K., Walsh, J., Cusick, M. E., Xia, Y., Barabasi, A. L., Iakoucheva, L. M., Aloy, P., De Las Rivas, J., Tavernier, J., Calderwood, M. A., Hill, D. E., Hao, T., Roth, F. P., and Vidal, M. (2014) *Cell* **159**, 1212-1226
118. Dherin, C., Gueneau, E., Francin, M., Nunez, M., Miron, S., Liberti, S. E., Rasmussen, L. J., Zinn-Justin, S., Gilquin, B., Charbonnier, J. B., and Boiteux, S. (2009) *Mol Cell Biol* **29**, 907-918
119. Langland, G., Kordich, J., Creaney, J., Goss, K. H., Lillard-Wetherell, K., Bebenek, K., Kunkel, T. A., and Groden, J. (2001) *J Biol Chem* **276**, 30031-30035
120. Pedrazzi, G., Perrera, C., Blaser, H., Kuster, P., Marra, G., Davies, S. L., Ryu, G. H., Freire, R., Hickson, I. D., Jiricny, J., and Stagljar, I. (2001) *Nucleic Acids Res* **29**, 4378-4386
121. Matton, N., Simonetti, J., and Williams, K. (2000) *J Biol Chem* **275**, 17808-17813

122. Schmutte, C., Marinescu, R. C., Sadoff, M. M., Guerrette, S., Overhauser, J., and Fishel, R. (1998) *Cancer Res* **58**, 4537-4542
123. Hill, S. J., Rolland, T., Adelmant, G., Xia, X., Owen, M. S., Dricot, A., Zack, T. I., Sahni, N., Jacob, Y., Hao, T., McKinney, K. M., Clark, A. P., Reyon, D., Tsai, S. Q., Joung, J. K., Beroukhim, R., Marto, J. A., Vidal, M., Gaudet, S., Hill, D. E., and Livingston, D. M. (2014) *Genes Dev* **28**, 1957-1975
124. MacKay, C., Declais, A. C., Lundin, C., Agostinho, A., Deans, A. J., MacArtney, T. J., Hofmann, K., Gartner, A., West, S. C., Helleday, T., Lilley, D. M., and Rouse, J. (2010) *Cell* **142**, 65-76
125. Meetei, A. R., Sechi, S., Wallisch, M., Yang, D., Young, M. K., Joenje, H., Hoatlin, M. E., and Wang, W. (2003) *Mol Cell Biol* **23**, 3417-3426
126. Smogorzewska, A., Desetty, R., Saito, T. T., Schlabach, M., Lach, F. P., Sowa, M. E., Clark, A. B., Kunkel, T. A., Harper, J. W., Colaiacovo, M. P., and Elledge, S. J. (2010) *Mol Cell* **39**, 36-47
127. Stehling, O., Vashisht, A. A., Mascarenhas, J., Jonsson, Z. O., Sharma, T., Netz, D. J., Pierik, A. J., Wohlschlegel, J. A., and Lill, R. (2012) *Science* **337**, 195-199
128. Roux, K. J., Kim, D. I., Raida, M., and Burke, B. (2012) *The Journal of cell biology* **196**, 801-810
129. Rhee, H. W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., and Ting, A. Y. (2013) *Science* **339**, 1328-1331
130. Ghodgaonkar, M. M., Kehl, P., Ventura, I., Hu, L., Bignami, M., and Jiricny, J. (2014) *Nature communications* **5**, 4990
131. Bartlett, J. M., and Stirling, D. (2003) *Methods Mol Biol* **226**, 3-6
132. Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K., and Ting, A. Y. (2015) *Nat Methods* **12**, 51-54
133. Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R., and Watts, J. D. (2009) *Nat Biotechnol* **27**, 378-386
134. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) *Electrophoresis* **20**, 3551-3567
135. Searle, B. C. (2010) *Proteomics* **10**, 1265-1269
136. Liu, G., Zhang, J., Larsen, B., Stark, C., Breitkreutz, A., Lin, Z. Y., Breitkreutz, B. J., Ding, Y., Colwill, K., Pasculescu, A., Pawson, T., Wrana, J. L., Nesvizhskii, A. I., Raught, B., Tyers, M., and Gingras, A. C. (2010) *Nat Biotechnol* **28**, 1015-1017
137. Eng, J. K., Jahan, T. A., and Hoopmann, M. R. (2013) *Proteomics* **13**, 22-24
138. Craig, R., and Beavis, R. C. (2004) *Bioinformatics* **20**, 1466-1467
139. Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J. K., Martin, D. B., Nesvizhskii, A. I., and Aebersold, R. (2010) *Proteomics* **10**, 1150-1159
140. Choi, H., Larsen, B., Lin, Z. Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z. S., Tyers, M., Gingras, A. C., and Nesvizhskii, A. I. (2011) *Nat Methods* **8**, 70-73
141. Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A. C., and Choi, H. (2014) *J Proteomics* **100**, 37-43
142. Brieger, A., Plotz, G., Hinrichsen, I., Passmann, S., Adam, R., and Zeuzem, S. (2012) *PLoS One* **7**, e31863

143. Shcherbakova, P. V., Hall, M. C., Lewis, M. S., Bennett, S. E., Martin, K. J., Bushel, P. R., Afshari, C. A., and Kunkel, T. A. (2001) *Mol Cell Biol* **21**, 940-951
144. Yoshizawa-Sugata, N., and Masai, H. (2009) *J Biol Chem* **284**, 20718-20728
145. Im, J. S., Ki, S. H., Farina, A., Jung, D. S., Hurwitz, J., and Lee, J. K. (2009) *Proc Natl Acad Sci U S A* **106**, 15628-15632
146. Hsieh, C. L., Lin, C. L., Liu, H., Chang, Y. J., Shih, C. J., Zhong, C. Z., Lee, S. C., and Tan, B. C. (2011) *Nucleic Acids Res* **39**, 4048-4062
147. Hao, J., de Renty, C., Li, Y., Xiao, H., Kemp, M. G., Han, Z., DePamphilis, M. L., and Zhu, W. (2015) *Embo J* **34**, 2096-2110
148. Chen, Z., Tran, M., Tang, M., Wang, W., Gong, Z., and Chen, J. (2016) *Mol Cell Proteomics*
149. http://www.ebi.ac.uk/intact/.
150. Wray, J., Liu, J., Nickoloff, J. A., and Shen, Z. (2008) *Cancer Res* **68**, 2699-2707
151. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., and Jensen, L. J. (2013) *Nucleic Acids Res* **41**, D808-815
152. Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J. H. (1992) *Cell* **71**, 939-953
153. Demers, C., Chaturvedi, C. P., Ranish, J. A., Juban, G., Lai, P., Morle, F., Aebersold, R., Dilworth, F. J., Groudine, M., and Brand, M. (2007) *Mol Cell* **27**, 573-584
154. Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y., and Fuks, F. (2006) *Nature* **439**, 871-874
155. Sander, J. D., and Joung, J. K. (2014) *Nat Biotechnol* **32**, 347-355
156. Hsu, P. D., Lander, E. S., and Zhang, F. (2014) *Cell* **157**, 1262-1278
157. Engel, J. D., and Dodgson, J. B. (1981) *Proc Natl Acad Sci U S A* **78**, 2856-2860
158. Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J., and Maltais, L. J. (2002) *Genomics* **80**, 487-498
159. Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013) *Cell* **154**, 442-451
160. Schultz, D. C., Friedman, J. R., and Rauscher, F. J., 3rd. (2001) *Genes Dev* **15**, 428-443
161. Frietze, S., O'Geen, H., Blahnik, K. R., Jin, V. X., and Farnham, P. J. (2010) *PLoS One* **5**, e15082
162. van de Vrugt, H. J., Eaton, L., Hanlon Newell, A., Al-Dhalimy, M., Liskay, R. M., Olson, S. B., and Grompe, M. (2009) *Cancer Res* **69**, 9431-9438
163. Xie, J., Guillemette, S., Peng, M., Gilbert, C., Buermeyer, A., and Cantor, S. B. (2010) *Cancer Prev Res (Phila)* **3**, 1409-1416
164. Brooks, P. J., and Theruvathu, J. A. (2005) *Alcohol* **35**, 187-193
165. Suchanek, M., Radzikowska, A., and Thiele, C. (2005) *Nat Methods* **2**, 261-267
166. Raschle, M., Smeenk, G., Hansen, R. K., Temu, T., Oka, Y., Hein, M. Y., Nagaraj, N., Long, D. T., Walter, J. C., Hofmann, K., Storchova, Z., Cox, J., Bekker-Jensen, S., Mailand, N., and Mann, M. (2015) *Science* **348**, 1253671

10 Acknowledgements

First of all, I want to thank Prof. Dr. Josef Jiricny. Starting already in the interview, Joe has been a great adviser regarding not only research-related questions but also other fields of life. He advised me to do my PhD in Zurich and also to join his laboratory and I believe it was a very good decision. I could follow my ideas under his supervision and explore many different techniques. He always has an open door and tries to find solutions together rather than telling me what to do next. He has always been very helpful and motivating when things did not go as I wished and he never gave me the feeling that failed experiments were my fault. He encouraged me to try new things and helped me to find collaborators for the mass spectrometry experiments. Thanks to the fact that he motivated me to go and to talk the people at ETH by my-self I found out that I can achieve a great deal in the scientific world without a powerful person around. I really like the team and the friendly atmosphere in our lab. I think that this is Joes merit as he knows how to pick the right people to keep the up freedom and the fun. Besides the actual research, Joe was always interested in my future and that of all the other lab members. It is a very nice feeling that he atually cares about our future. Although, he would like all of us to continue with academic research, he has never put any obstacles in peoples way but even helped them to find their way outside academia. I never regretted moving to Zurich and joining Joes lab. I am very grateful that he gave me the opportunity to do a PhD in his laboratory.

Secondly, I would like to thank my committee members Dr. Matthias Gstaiger and Prof. Dr. Petr Cjeka. I am very grateful that they agreed to join my committee and actively contributed to the development of my projects. It was very easy to communicate with them and to identify the important aims of my projects together. It was great to discuss the new techniques for LC-MS such as BioID and APEX and I would like to thank Matthias for helping me with the APEX approach and for supporting me to start a collaboration with Emanuela Milani from the

Wollscheid lab. It was an excellent experience to work with her and to find solutions for the APEX project. It is thanks to her that I now understand the basics of the MS analyses and the statistics. I also want to thank Peter Gehrig of the Funtional Genomics Center Zurich for his assistance with the GFP-MLH1 MS data.

A big word of thanks goes to my colleagues in the lab. I think we are a great team and had a very nice time together during the last four years. We could not only discuss our science-related problems and tried to help each other but also became very close friends and did and still do a lot of thinks togehter. We did so many things outside the lab like our skiweekends and our trip on the Aare and the lunch and coffe breaks are allways something to look forward, when the experiments do not work out as planned. What I like a lot is that we can always share and talk about problems that come along during a PhD. So thank you so much Tine, Lia, Julia, Anja, Svenja, Simone, Sascha, Saho, Mariela, Rachel, Sara, Antonio, Shun, Steffi, Medini and Maite.

I want to thank Mariela again in a separate section for all the work she has invested in my projects. Starting with making the MMR substrates, showing and helping me to make nuclear extracts and finally performing the MMR assays. Without her those experiments would not have been possible. I am really grateful for this.

I also want to thank the PhD Program Cancer Biology and the Life Science Graduate Program, which made my PhD life more easy and provided great opportunities to make friends and meet fellow PhD students.

Als nächstes möchte ich meiner Familie danken, die mich in all den Jahren meines Studiums nicht nur finanziell unterstützt hat. Meine Eltern waren immer für mich da, besonders dann, wenn ich sie am meisten gebraucht habe. Ich fand es immer toll, dass sie mich an allen Orten meiner Studienzeit und während des Doktors mehrmals pro Jahr besucht haben und wir so zusammen die Städte

Rostock, Heidelberg, Penzberg und Zürich kennenlernen konnten. Ich freue mich riesig, dass die ganze Familie zu meiner Verteidigung nach Zürich kommt.

Zum Ende geht ein riesiges und herzliches Dankeschön an Leonie. Seit 6 Jahren hat sie mit mir zusammen alle Höhen und Tiefen des Studiums und der Doktorzeit erlebt und einen sehr großen Teil dazu beigetragen, dass ich es bis hierhin geschafft habe. Ich weis gar nicht mehr wieviele Präsentationen und Abstracts sie für mich korrigiert hat. Und natürlich hat sie meine Masterarbeit gegengelesen und nun auch die Doktorarbeit. Dank ihr konnte ich diese mit bestmöglichem Englisch und viel weniger Rechtschreibfehlern an Joe schicken. Vielen Dank, dass du mir dabei geholfen hast und mich immer unterstützt hast.

11 Curriculum Vitae

General Information

First name:	Martin
Last name:	Falke
Date of birth:	February 18, 1986
Place of birth:	Schwedt/Oder, Germany
Nationality:	German
Address:	Keltenstrasse 42 8044 Zurich Switzerland
Phone:	+41 44 635 34 79
Mobile:	+41 78 634 33 40
E-mail:	falke@imcr.uzh.ch

Education and Training

07/2012 - present	**ETH Zurich,** Graduate student with Prof. Dr. Jiricny Cancer Biology PhD Program Zurich Main project: "Second Generation Protein Replacement System to Study MLH1 Missense Mutations"
10/2009 - 04/2012:	**University of Heidelberg**, Master studies "Molecular Biosciences", Major: Cancer Biology (04/2011: M.Sc. degree in Molecular Bioscience, grade 1.2; Master thesis “Regulation of CAMTA1, a Newly Described Tumor Suppressor Gene”)
10/2006 - 09/2009:	**University of Rostock**, Bachelor studies in Medical Biotechnology (09/2009: B.Sc. degree in Medical Biotechnology, grade: 1.4; Bachelor

	thesis "Generation of a Microglia-Specific Expression Plasmid")
07/2005 - 03/2006:	**Military service**
06/2006:	German **A-level** (Abitur), grade: 1.5

Publications

Loss of lamin B1 results in prolongation of S phase and decondensation of chromosome territories. Camps, J., Wangsa, D., **Falke, M**., Brown, M., Case, C. M., Erdos, M. R., and Ried, T. (2014) *Faseb J* **28**, 3423-3434

Second generation protein replacement system to study MLH1 missense mutations. **Falke, M.,** Artola Borán, M., Kapaklikaya, E., Jiricny, J. (Manuscript in preparation)

Practical Experience

10/2011 – 04/2012	Master thesis at the **German Cancer Research Center (Deutsches Krebsforschungs-zentrum, DKFZ)** Heidelberg, Germany; Tumorgenetics, Supervision by Dr. Frank Westermann and Dr. Kai-Oliver Henrich
05/2011 – 08/2011:	Internship at **Roche Diagnostics GmbH** Penzberg, Germany; Pharma Research and Early Development; Laboratory of Dr. Claudio Sustmann
01/2011 – 03/2011:	Internship at the **German Cancer Research Center (Deutsches Krebsforschungs-zentrum, DKFZ)** Heidelberg, Germany; Molecular RNA Biology and Cancer, Laboratory of Dr. Sven Diederichs

08/2010 - 01/2011:	Internship at the **National Institutes of Health/National Cancer Institute** in Bethesda, USA; Laboratory of Dr. Thomas Ried
10/2008 - 06/2009:	Student assistant at the **University of Rostock**, Clinic for Neurology, Neurodegeneration Research Lab; Laboratory of Prof. Jens Pahnke
08/2008 - 09/2008:	Internship at the **University of Rostock**, Clinic for Neurology, Neurodegeneration Research Lab; Laboratory of Prof. Jens Pahnke
04/2006 - 07/2006:	Internship at the **Hospital "Klinikum Südstadt"** in Rostock, Germany

Additional Skills and Qualifications

Language skills:	**German** (native speaker) **English** (fluent) **Russian** (basic knowledge)
IT:	Excellent knowledge of the **MS Office** programs